

Laporan Penelitian Lanjut
Bidang Ilmu



*Kesesuaian Metode Regresi Nonparametrik Spline,
B-spline, dan P-spline
dalam Menduga Kurva Regresi*

Oleh:
Dra. Harmi Sugiarti, M.Si

Pusat Penelitian dan Pengabdian kepada Masyarakat
UNIVERSITAS TERBUKA
2012

LEMBAR PENGESAHAN

Laporan Penelitian Lembaga Penelitian dan Pengabdian kepada Masyarakat UT

1. a. Judul Penelitian : Kesesuaian Metode Regresi Nonparametrik
Spline, B-spline dan *P-spline* dalam Menduga
Kurva Regresi
b. Bidang Penelitian : Bidang Ilmu
c. Klasifikasi Penelitian : Penelitian Lanjut
d. Bidang Ilmu : Statistika
2. Ketua Peneliti
a. Nama : Dra. Harmi Sugiarti, M.Si
b. NIP : 19670311 199202 2 001
c. Golongan Kepangkatan : III/d
d. Jabatan Akademik : Lektor
e. Fakultas/ Unit Kerja : FMIPA
3. Anggota Tim Peneliti
a. Jumlah Anggota : -
b. Nama/Unit Kerja : -
4. Lama Penelitian : 9 (sembilan) bulan
5. Biaya Penelitian : Rp. 30.000.000,- (tiga puluh juta rupiah)
6. Sumber Biaya : LPPM UT

Pondok Cabe, 13 Desember 2012

Mengetahui,
Dekan FMIPA-UT

Ketua Peneliti,

Dr. Nuraini Soleiman, M.Ed
NIP. 19540730 198601 2 001

Dra. Harmi Sugiarti, M.Si
NIP. 19670311 199202 2 001

Mengetahui,
Ketua LPPM- UT

Menyetujui,
Kepala Pusat Keilmuan LPPM-UT

Dra. Dewi A. Padmo Putri, MA. Ph.D
NIP. 19610724 198710 2 001

Dra. Endang Nugraheni, M.Ed, M.Si
NIP. 19570422 198503 2 001

RINGKASAN

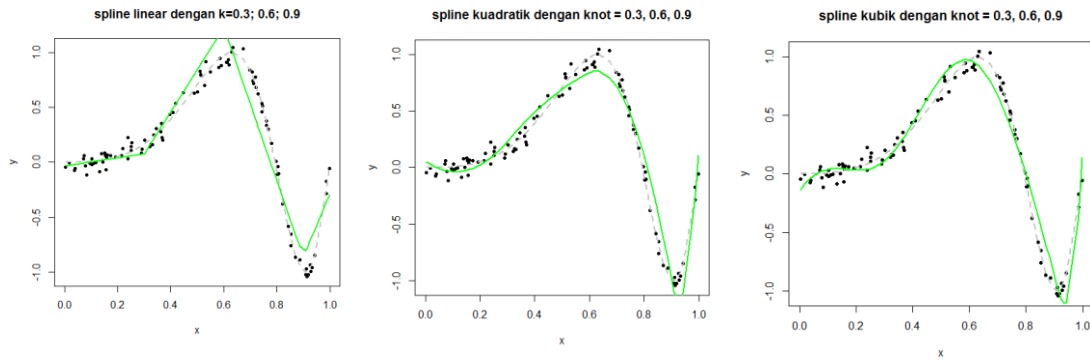
Analisis regresi merupakan salah satu analisis dalam statistika yang sangat sering digunakan untuk memperlihatkan hubungan dan pengaruh variabel prediktor (variabel bebas) terhadap variabel respon (variabel tak bebas). Jika model yang diasumsikan benar, maka pendugaan parametrik sangat efisien. Namun, jika model yang diasumsikan tidak benar, maka dapat menyebabkan interpretasi data yang menyesatkan. Selain itu, model parametrik juga mempunyai keterbatasan untuk menduga pergerakan data yang tidak diharapkan, sehingga diperlukan pendugaan yang bersifat nonparametrik.

Metode regresi nonparametrik yang dapat digunakan untuk menduga kurva regresi antara lain dengan menggunakan pendekatan regresi spline. Regresi spline adalah pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. Spline adalah penduga yang diperoleh dengan meminimumkan *penalized least square*, yaitu kriteria pendugaan yang menggabungkan *goodness-of-fit* dengan kemulusan kurva, dimana kedua ukuran ini diatur oleh suatu parameter pemulusan. Basis fungsi yang digunakan dalam pendekatan regresi spline adalah *truncated* dan *B-spline*. Pemilihan model optimal (terbaik) menggunakan kriteria *Mean Squared Error (MSE)* atau *Generalized Cross Validation (GCV)* dan taksiran parameter menggunakan metode kuadrat terkecil.

Penelitian ini bertujuan untuk mengkaji dan mengevaluasi penduga pada kurva regresi linear, kuadratik, maupun kubik dengan pendekatan model spline, B-spline, dan P-spline pada data simulasi dan data terapan. Data simulasi dibangkitkan dan diolah dengan bantuan program R versi 2.14.1. Data terapan diperoleh dari SRS-UT berupa Indek Prestasi (IP) pada semester satu dan lama studi mahasiswa PS Statistika FMIPA Universitas Terbuka yang lulus pada masa ujian 1989.2 sampai dengan 2011.2.

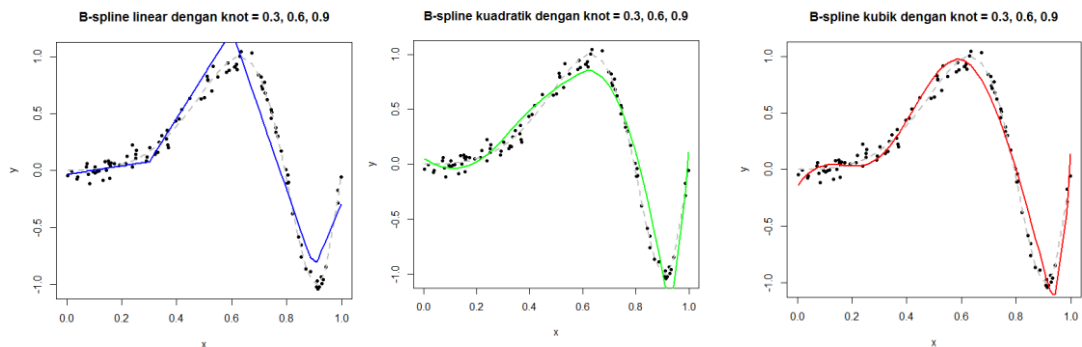
Data bangkitan sebanyak 100 pengamatan berpasangan menunjukkan bahwa pendekatan model spline dengan basis truncated memberikan performa yang baik dalam memodelkan data baik untuk penduga kurva spline linear, kuadratik, dan kubik pada knot $k_1 = 1$, $k_2 = 2$, $k_3 = 3$. Sebagaimana ditunjukkan

pada Gambar 1, tampak bahwa performa taksiran fungsi spline linear, kuadratik, dan kubik pada knot $k_1 = 1$, $k_2 = 2$, $k_3 = 3$ mendekati garis sesungguhnya (garis putus-putus).



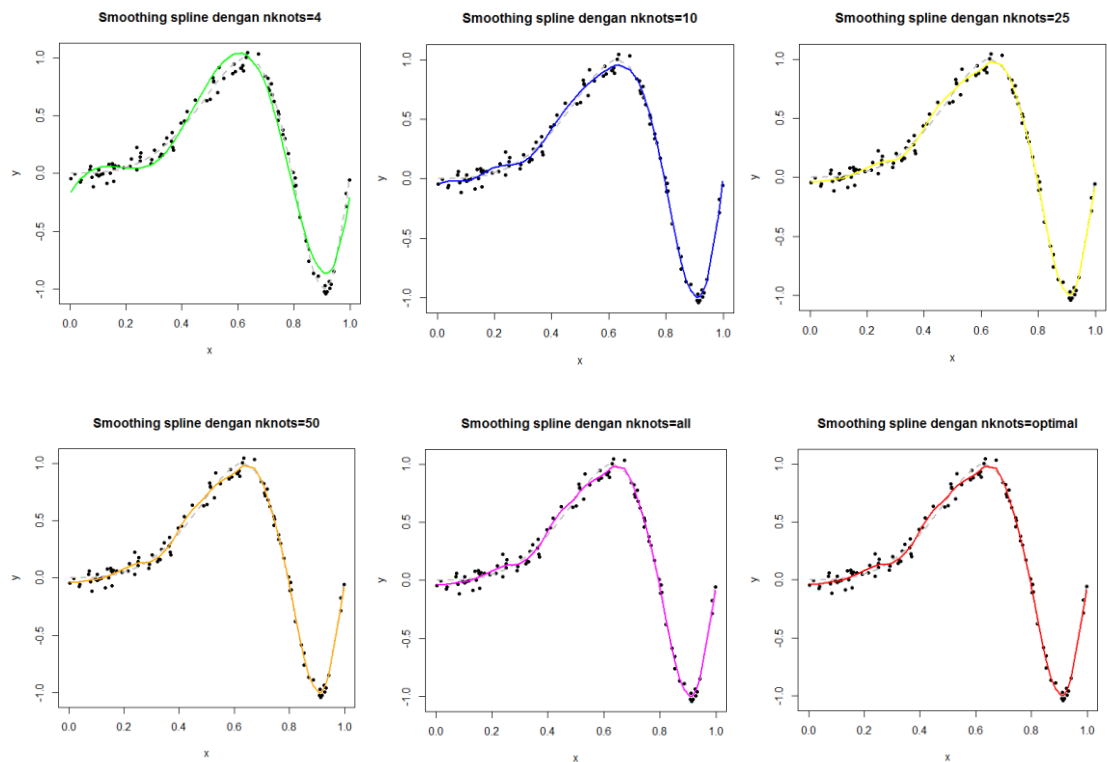
Gambar 1. Kurva Spline Data Simulasi

Pendekatan model spline dengan basis B-spline menunjukkan bahwa penduga kurva B-spline linear, kuadratik, dan kubik pada knot $k_1 = 1$, $k_2 = 2$, $k_3 = 3$ mempunyai performa yang baik dalam memodelkan data seperti ditunjukkan pada Gambar 2.



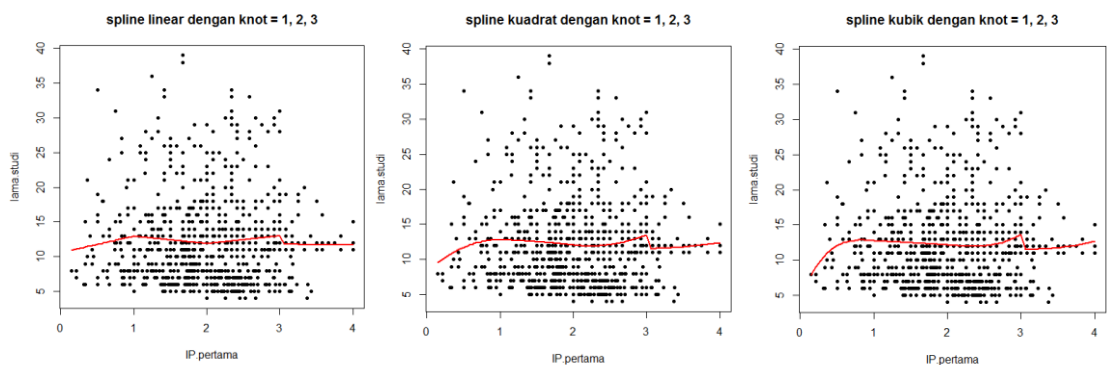
Gambar 2. Kurva B-spline Data Simulasi

Pendekatan model smoothing spline dengan basis B-spline menunjukkan bahwa penduga kurva P-spline dengan menggunakan sebanyak 62 knot merupakan model dengan performa yang optimal seperti ditunjukkan pada Gambar 3.



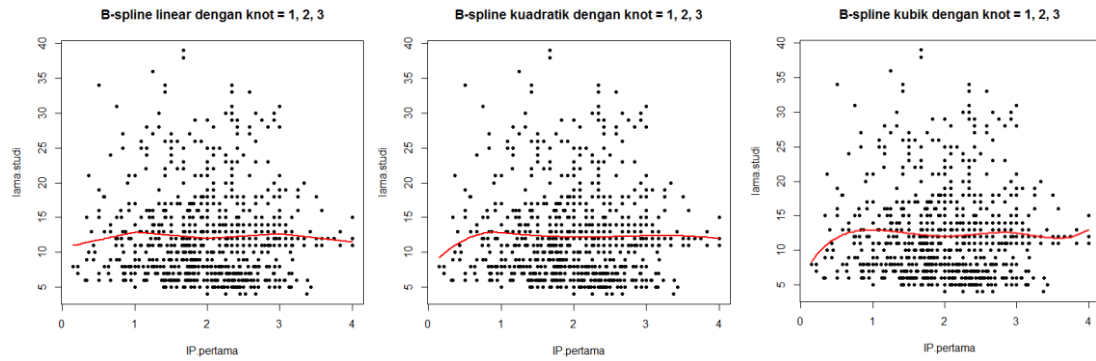
Gambar 3. Kurva P-Spline Data Simulasi

Berdasarkan data terapan sebanyak 835 mahasiswa Program Studi Statistika FMIPA Universitas Terbuka yang lulus pada masa ujian 1989.2 sampai dengan 2011.2, pendekatan model spline dengan basis truncated pada data terapan menunjukkan bahwa penduga kurva spline linear, kuadratik, dan kubik pada berbagai knot mempunyai performa yang kurang baik dalam memodelkan data seperti disajikan pada Gambar 4.



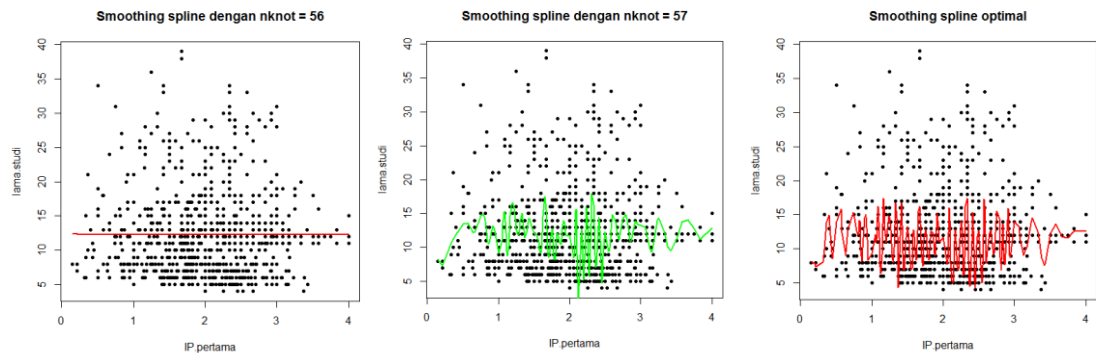
Gambar 4. Kurva Spline Data Terapan

Pendekatan model spline dengan basis B-spline untuk data terapan menunjukkan bahwa penduga kurva B-spline linear, kuadratik, dan kubik pada berbagai knot mempunyai performa yang kurang baik dalam memodelkan data seperti disajikan pada Gambar 5.



Gambar 5. Kurva B-spline Data Terapan

Performa model P-spline menunjukkan bahwa penduga kurva P-spline dengan menggunakan berbagai jumlah knot mempunyai performa yang kurang baik dalam memodelkan data seperti disajikan pada Gambar 6.



Gambar 6. Kurva P-spline Data Terapan

Kesesuaian Metode Regresi Nonparametrik Spline, B-spline dan P-spline dalam Menduga Kurva

Abstract

Regression modeling with nonparametric regression method is required as an alternative method of standard methods (parametric regression method) to obtain a more accurate model when there is assumption "true model" violation. There are several nonparametric regression methods that can be used, including the spline methods. This study aims to assess performance of the spline basis method with truncated (spline method) , spline method with B-spline basis (B-spline), and penalized spline method with B-spline basis (P-spline). This study used simulated data and applied data. Simulated data are generated and processed with the R programs version 2.14.1. Applied data are obtained from the SRS-UT with the first grade point and life time study for Statistics Department students in 1989.2-2011.2 periods. Base on the 100 paired observations from generated data and the generalized cross validation (GCV) and mean squared error (MSE) criterion , the spline method and B-spline method with three knots have good performance to fit the data, while the P-spline method have optimal performance when 10 knots are used. Base on 835 paired observations from real data, there is not significantly different performance of the spline method, B-spline method, and the P-spline method.

Keywords: spline methods, nonparametric regression

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	i
RINGKASAN	ii
ABSTRAK	vi
DAFTAR ISI	vii
BAB I. PENDAHULUAN	
A. Latar Belakang	1
B. Perumusan Masalah	3
C. Tujuan Penelitian	3
D. Manfaat Penelitian	3
BAB II. TINJAUAN PUSTAKA	
A. Regresi Nonparametrik <i>Spline</i>	5
B. Regresi Nonparametrik <i>B-Spline</i>	7
C. Regresi Nonparametrik <i>P-Spline</i>	15
D. Kriteria Asimtotik	18
BAB III. METODE PENELITIAN	
A. Desain Penelitian	19
B. Data	19
C. Tahapan Analisis	20
BAB IV. HASIL DAN PEMBAHASAN	
A. Hasil Simulasi	22
B. Hasil Terapan	38
BAB V. KESIMPULAN DAN SARAN	
DAFTAR PUSTAKA	
BIODATA	

BAB I

PENDAHULUAN

A. LATAR BELAKANG

Analisis regresi merupakan salah satu analisis dalam statistika yang sangat sering digunakan untuk memperlihatkan hubungan dan pengaruh variabel prediktor (variabel bebas) terhadap variabel respon (variabel tak bebas). Misalkan y adalah variabel respon dan x adalah variabel prediktor, secara umum hubungan antara x dan y untuk n pengamatan dapat ditulis sebagai $y_i = f(x_i) + \varepsilon_i$, $i = 1, 2, \dots, n$ dengan ε adalah sesatan random dan $f(x_i)$ merupakan kurva regresi. Pendekatan yang sering digunakan untuk menduga kurva regresi $f(x_i)$ adalah pendekatan parametrik, dimana diasumsikan bahwa bentuk fungsi dari $f(x_i)$ dapat digambarkan secara lengkap oleh satu set parameter. Pada pendekatan parametrik, kurva regresi dapat diwakili oleh model parametrik atau bias dari taksirannya berupa nilai yang dapat diabaikan (Hardle, 1990). Jika model yang diasumsikan benar, maka pendugaan parametrik sangat efisien. Jika model yang diasumsikan tidak benar, maka dapat menyebabkan interpretasi data yang menyesatkan. Selain itu, model parametrik juga mempunyai keterbatasan untuk menduga pergerakan data yang tidak diharapkan.

Beberapa kajian telah dilakukan untuk mengetahui tingkat efisiensi metode parametrik, diantaranya adalah tingkat efisiensi metode M, metode LMS, dan metode OLS pada regresi robust. Hasil kajian untuk data yang tidak memenuhi asumsi, khususnya data mengandung pencilan, menunjukkan bahwa metode LMS memberikan nilai koefisien determinasi yang lebih konsisten dibanding metode OLS dan metode M (Sugiarti & Megawarni, 2010).

Selain itu, kajian juga dilakukan untuk mengetahui kinerja metode LTS sebagai metode alternatif dibanding metode OLS, metode M, dan metode LMS dalam menaksir koefisien garis regresi jika data mengandung pencilan. Hasil kajian menunjukkan bahwa pada data yang mengandung pencilan, metode LTS mempunyai kinerja yang tidak jauh berbeda dengan metode robust lainnya (metode M dan metode LMS) dalam menaksir koefisien garis regresi (Sugiarti & Megawarni, 2011).

Jika dilihat dari hasil kajian yang telah dilakukan, metode robust yang digunakan dalam model parametrik cukup baik untuk digunakan. Namun, penggunaan metode robust pada data terapan/real masih dianggap kurang menjelaskan hubungan yang ada diantara peubah-peubah. Hal ini mungkin disebabkan karena pola hubungan yang ada dalam model tidak dapat dijelaskan hanya dengan asumsi hubungan linear, sementara hubungan kuadratik dan kubik tidak dijelaskan oleh model yang ada, dengan kata lain informasi tentang bentuk $f(x_i)$ tidak ada. Tidak adanya informasi tentang bentuk $f(x_i)$ mengakibatkan penggunaan model parametrik dianggap kurang pas dan dianggap perlu untuk melakukan kajian dengan pendekatan nonparametrik.

Sebagaimana disebutkan dalam Eubank (1999), jika tidak ada informasi apapun tentang bentuk $f(x_i)$, maka dapat digunakan pendekatan nonparametrik. Pendekatan nonparametrik dianggap lebih fleksibel dibanding pendekatan parametrik, karena pendekatan nonparametrik tidak tergantung pada asumsi bentuk kurva tertentu, dalam hal ini $f(x_i)$ diasumsikan termuat dalam ruang fungsi berdimensi tak berhingga. Selain itu, pada pendekatan regresi nonparametrik, data diharapkan dapat mencari bentuk taksirannya sendiri.

Metode regresi nonparametrik yang dapat digunakan untuk menduga kurva regresi antara lain dengan menggunakan pendekatan regresi spline. Regresi spline adalah pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. Spline adalah penduga yang diperoleh dengan meminimumkan *penalized least square*, yaitu kriteria pendugaan yang menggabungkan *goodness-of-fit* dengan kemulusan kurva, dimana kedua ukuran ini diatur oleh suatu parameter pemulusan.

Wahba (1990) menunjukkan bahwa spline memiliki sifat-sifat statistik yang berguna dan perlu dipertimbangkan sebagai satu metode untuk menganalisis hubungan regresi. Basis fungsi yang digunakan dalam pendekatan regresi spline adalah *truncated* dan *B-spline*. Pemilihan model optimal (terbaik) menggunakan kriteria *Mean Squared Error (MSE)* atau *Generalized Cross Validation (GCV)* dan taksiran parameter menggunakan metode kuadrat terkecil.

Beberapa kajian tentang model nonparametrik telah dilakukan, diantaranya penggunaan B-spline dalam memodelkan nilai ujian masuk terhadap IPK mahasiswa Jurusan Desain Komunikasi Visual Universitas Kristen Petra Surabaya (Budiantara, I.N. dkk, 2006). Selain itu, kajian tentang regresi nonparametrik juga dilakukan pada spline dengan menggunakan berbagai faktor pemulus (Lee, 2004) dan (Adisantoso, 2010). Dalam penelitian ini akan dikaji penggunaan regresi spline dalam memodelkan IP semester pertama terhadap lama studi mahasiswa Program Studi Statistika FMIPA Universitas Terbuka.

B. PERUMUSAN MASALAH

Berdasarkan uraian di atas, permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana menentukan penduga pada kurva regresi linear, kuadratik, maupun kubik dengan pendekatan model spline, B-spline, dan P-spline.
2. Bagaimana mengevaluasi kesesuaian kurva regresi dengan pendekatan model spline B-spline, dan P-spline untuk data simulasi dan data terapan.

C. TUJUAN PENELITIAN

Berdasarkan perumusan masalah diatas, penelitian ini bertujuan untuk:

1. Mengkaji penduga pada kurva regresi dengan pendekatan model spline, B-spline, dan P-spline.
2. Mengevaluasi kesesuaian model spline, B-spline, dan P-spline untuk data simulasi dan data terapan.

D. MANFAAT PENELITIAN

Hasil penelitian ini diharapkan dapat memberikan masukan yang lebih jelas bagi praktisi atau pengguna statistik tentang spline, B-spline, dan P-spline sehingga dapat menentukan penggunaan pendekatan spline, B-spline, dan P-spline dengan tepat. Manfaat lain dari penerapan metode pada data terapan diharapkan dapat digunakan sebagai rujukan dalam penelitian yang sesuai serta dapat digunakan sebagai dasar pengambilan kebijakan dalam pelayanan

kemahasiswaan dan peningkatan prestasi akademik mahasiswa FMIPA
Universitas Terbuka.

BAB II

TINJAUAN PUSTAKA

A. REGRESI NONPARAMETRIK SPLINE

Jika diamati nilai-nilai variabel respon Y dan variabel bebas X yang telah ditentukan, diperoleh hasil pengamatan berpasangan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ secara umum hubungan antara X dan Y untuk n pengamatan dapat ditulis sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

dengan ε adalah sesatan random yang bersifat identik, independen dengan nilai tengah nol dan variansi sama dengan σ^2 ; fungsi $f(x_i)$ adalah kurva regresi yang diasumsikan tidak diketahui bentuknya dan mulus (*smooth*) dalam arti kontinu dan differensiabel, dengan pemilihan ruang fungsi berdasarkan sifat kemulusan pada fungsi $f(x_i)$ tersebut.

Fungsi dikatakan *smooth* secara geometrik jika gradiennya berubah tidak terlalu cepat, sehingga dapat menggunakan suatu titik yang berada disekitar titik tersebut sebagai taksirannya. Berdasarkan konsep tersebut, misalkan f adalah fungsi *smooth* dalam arti f termuat dalam suatu ruang fungsi Sobolev $W_2^m[a, b] = \{f / f, f', f'', \dots, f^{(m-1)} \text{ kontinu absolut pada } f^{(m)} \in L_2[a, b]\}$, dengan $f^{(m)}$ adalah fungsi turunan ke m dan $L_2[a, b]$ adalah himpunan fungsi yang kuadratnya terintegral pada interval $[a, b]$ atau dapat dinyatakan sebagai:

$$L_2[a, b] = \left\{ f \left| \int_a^b [f^{(m)}(x)]^2 dt < \infty \right. \right\} \quad (\text{Eubank, 1999}).$$

Pendekatan spline kurva dibentuk berdasarkan titik-titik knot yang kemudian saling dihubungkan. Kurva pendekatannya bisa kurva interpolasi polinomial, kurva *Bezier*, kurva spline (Lyche dan Morken, 2008). Spline merupakan salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen, dimana segmen-segmen polinomial yang berbeda digabungkan bersama pada knot-knot dengan suatu cara yang menjamin sifat kontinuitas tertentu. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa,

sehingga memungkinkannya untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal yaitu suatu fungsi atau data. Spline didefinisikan sebagai polinomial tersegmen (*piecewise polinomial*) yang memiliki sejumlah batasan kontinuitas yang maksimum sehingga spline merupakan polinomial tersegmen termulus yang mungkin saja masih mempunyai sifat tersegmen.

Fungsi spline berorde m dengan satu variabel bebas dan basis *truncated* adalah sembarang fungsi yang secara umum dapat disajikan dalam bentuk:

$$E(y) = f(x) = \beta_0 + \sum_{r=1}^{m-1} \beta_r x^r + \sum_{j=1}^K \beta_{m-1+j} (x - k_j)_+^{m-1} \quad (2)$$

dengan basis *truncated* yakni:

$$(x - k_j)_+^{m-1} = \begin{cases} (x - k_j)^{m-1} & ; x \geq k_j \\ 0 & ; x < k_j \end{cases} \quad (3)$$

dengan k_j adalah knot ke j dari variabel x , $j = 1, 2, 3, \dots, K$ dan K banyaknya knot (Eubank, 1999). Knot adalah suatu titik fokus dalam fungsi spline sedemikian sehingga kurva yang dibentuk tersegmen pada titik tersebut. Spline orde m dengan knot pada nilai-nilai tertentu misalnya disebut sebagai fungsi f , mempunyai sifat-sifat sebagai berikut:

- (1) Fungsi f adalah suatu polinomial tersegmen dengan orde m pada sub interval antar knot yang berurutan $[k_j, k_{j+1}]$.
- (2) Fungsi f memiliki sejumlah $m-2$ turunan yang kontinu.
- (3) Fungsi f memiliki turunan ke $m-1$ yang merupakan fungsi step dengan lompatan pada masing-masing knot yang berurutan k_1, k_2, \dots, k_K .
- (4) Nilai penduga parameter β diperoleh dengan metode ordinary least square (OLS), yaitu:

$$\hat{\beta} = [X'X]^{-1} [X'Y] \quad (4)$$

dengan

$$X = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \dots & x_{11}^{m-1} & (x_{11} - k_1)_+^{m-1} & (x_{11} - k_2)_+^{m-1} & \dots & (x_{11} - k_K)_+^{m-1} \\ 1 & x_{21} & x_{21}^2 & \dots & x_{21}^{m-1} & (x_{21} - k_1)_+^{m-1} & (x_{21} - k_2)_+^{m-1} & \dots & (x_{21} - k_K)_+^{m-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n1}^2 & \dots & x_{n1}^{m-1} & (x_{n1} - k_1)_+^{m-1} & (x_{n1} - k_2)_+^{m-1} & \dots & (x_{n1} - k_K)_+^{m-1} \end{bmatrix}$$

Fungsi basis *truncated spline* mempunyai kelemahan pada saat orde spline makin tinggi dan vektor knot yang makin besar, yakni akan membentuk matriks dalam perhitungan yang hampir singular, sehingga kemungkinan ada sebuah persamaan spline yang tidak dapat diselesaikan.

Dalam Eubank (1999) disebutkan bahwa salah satu ukuran performa atas penduga kurva regresi yang dapat digunakan adalah *mean squared error (MSE)* yang didefinisikan sebagai:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (5)$$

B. REGRESI NONPARAMETRIK B-SPLINE

Fungsi basis lain yang dapat digunakan dan kondisinya dianggap lebih baik dari basis *truncated* adalah basis B-spline (Eubank, 1999). Baik dan tidaknya sebuah model regresi nonparametrik, sangat dipengaruhi oleh fungsi basis yang digunakan untuk pendekatan. Beberapa penelitian banyak menggunakan basis *truncated* dan B-spline untuk menyelesaikan masalah. Botella dan Shariff (2003) menekankan sifat-sifat B-spline, misalnya adanya algoritma untuk mengevaluasi B-spline yang lebih efisien, serta B-spline mempunyai sifat ringkas.

Lyche dan Morken (2008) menyatakan B-spline mempunyai sifat antara lain *local knots*, *local support*, *positivity*, *piecewise polynomial*, *special values*, dan *smoothness*. Jika d menyatakan derajat polinomial nonnegatif dan $k = (k_j)$ adalah knot yang berurut, maka B-spline pada k mempunyai sifat:

- (1) *local knots*, yakni $B_{j,d}$ hanya bergantung pada knot $k_j, k_{j+1}, k_{j+2}, \dots, k_{j+d+1}$.
- (2) *local support*, jika ada x berada di luar interval $[k_j, k_{j+d+1})$ maka $B_{j,d}(x) = 0$ khususnya jika $k_j = k_{j+d+1}$ maka $B_{j,d}(x) = 0$.
- (3) *positivity*, jika $x \in (k_j, k_{j+d+1})$ maka $B_{j,d}(x) > 0$ dan interval tertutup $[k_j, k_{j+d+1}]$ dikatakan *support* dari $B_{j,d}$.
- (4) *piecewise polynomial*, B-Spline $B_{j,d}$ dapat ditulis sebagai:

$$B_{j,d}(x) = \sum_{k=j}^{j+d} B_{j,d}^k(x) B_{k,0}(x) \text{ dengan setiap } B_{j,d}^k(x) \text{ adalah polinomial derajat } d$$

- (5) *special values*, jika $z = k_{j+1} = k_{j+2} = \dots = k_{j+d} < k_{j+d+1}$ maka $B_{j,d}(z) = 1$ dan $B_{i,d}(z) = 0$ untuk $i \neq j$
- (6) *smoothness*, jika terdapat z sebanyak t kali diantara $k_j, k_{j+1}, k_{j+2}, \dots, k_{j+d+1}$ maka derivatif ke $0, 1, \dots, d-t$ dari $B_{j,d}$ akan kontinu pada z .

$$B_{j,d}(x) = \begin{cases} 0 & , \text{ jika } k_j = k_{j+1+d} \\ s_1(x) & , \text{ jika } k_j < k_{j+d} \text{ dan } k_{j+1} = k_{j+1+d} \\ s_2(x) & , \text{ jika } k_j = k_{j+d} \text{ dan } k_{j+1} < k_{j+1+d} \\ s_1(x) + s_2(x) & , \text{ untuk yang lainnya} \end{cases}$$

$$s_1(x) = \frac{x - k_j}{k_{j+d} - k_j} B_{j,d-1}(x)$$

$$s_2(x) = \frac{k_{j+1+d} - x}{k_{j+1+d} - k_{j+1}} B_{j+1,d-1}(x)$$

$$B_{j,0}(x) = \begin{cases} 1 & , x \in [k_j, k_{j+1}) \\ 0 & , x \notin [k_j, k_{j+1}) \end{cases}$$

Dalam Budiantara (2006) disebutkan bahwa kurva regresi berorde m dan basis B-Spline dengan knot pada k_1, k_2, \dots, k_K secara umum dapat ditulis dalam bentuk:

$$f(x) = \sum_{j=1}^{K+m} \beta_j B_{j-m,m}(x) \quad (6)$$

dengan $B_{j-m,m}(x)$ merupakan basis B-spline. Fungsi B-spline secara rekursif didefinisikan sebagai:

$$B_{j,m}(x) = \frac{x - k_j}{k_{j+m-1} - k_j} B_{j,m-1}(x) + \frac{k_{j+m} - x}{k_{j+m} - k_{j+1}} B_{j+1,m-1}(x) \quad (7)$$

$$\text{dengan } B_{j,1}(x) = \begin{cases} 1 & , x \in [k_j, k_{j+1}) \\ 0 & , \text{ untuk } x \text{ yang lainnya} \end{cases}$$

Guna membangun fungsi B-spline orde m , diperlukan tambahan knot sebanyak $2m$ yaitu: $k_{-(m-1)}, k_{-(m-2)}, \dots, k_{-1}, k_0, k_1, k_2, \dots, k_K, k_{K+1}, \dots, k_{K+m}$ dengan nilai $k_{-(m-1)} = k_{-(m-2)} = \dots = k_{-1} = k_0 = a$ dan $k_{K+1} = k_{K+2} = \dots = k_{K+m} = b$, biasanya a adalah minimum nilai x dan b adalah maksimum nilai x . Pada fungsi basis B-

spline linear ($m=2$) dengan K knot, diperlukan tambahan $2m=4$ knot yaitu $k_{-1}=k_0=a$, $k_{K+1}=k_{K+2}=b$, dan banyaknya knot menjadi $K+2$.

Koefisien $\beta_1, \beta_2, \dots, \beta_{m+K}$ pada kurva regresi $f(x) = \sum_{j=1}^{K+m} \beta_j B_{j-m,m}(x)$ dapat ditaksir dengan metode OLS, yaitu meminimumkan jumlah kuadrat sisaan:

$$e^2 = \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (8)$$

sehingga diperoleh:

$$\hat{\beta} = [B(x)' B(x)]^{-1} [B(x)' Y] \quad (9)$$

dengan $B(x)$ adalah matriks berukuran $n \times (m+K)$

$$B(x) = \begin{pmatrix} B_{1-m,m}(x_1) & B_{2-m,m}(x_1) & \cdots & B_{K,m}(x_1) \\ B_{1-m,m}(x_2) & B_{2-m,m}(x_2) & \cdots & B_{K,m}(x_2) \\ \vdots & \vdots & & \vdots \\ B_{1-m,m}(x_n) & B_{2-m,m}(x_n) & \cdots & B_{K,m}(x_n) \end{pmatrix} \quad (10)$$

Sebagai contoh, fungsi basis spline linear ($m=2$) dengan $K=1$ knot pada $k_1=2$ dan nilai $x \in [0,4]$ dapat dibangun sebagai berikut. Fungsi basis B-spline linear memerlukan tambahan $2m=4$ knot yaitu $k_{-1}=k_0=0$ dan $k_2=k_3=4$, sehingga knot menjadi $k_{-1}=k_0=0, k_1=2, k_2=k_3=4$. Selanjutnya matriks $B(x)$ berukuran $n \times 3$ akan dibentuk sebagai:

$$B(x) = [B_{-1,2}(x) \quad B_{0,2}(x) \quad B_{1,2}(x)] = \begin{pmatrix} B_{-1,2}(x_1) & B_{0,2}(x_1) & B_{1,2}(x_1) \\ B_{-1,2}(x_2) & B_{0,2}(x_2) & B_{1,2}(x_2) \\ \vdots & \vdots & \vdots \\ B_{-1,2}(x_n) & B_{0,2}(x_n) & B_{1,2}(x_n) \end{pmatrix}$$

$$\text{Basis: } B_{-1,2}(x) = \frac{x-k_{-1}}{k_0-k_{-1}} B_{-1,1}(x) + \frac{k_1-x}{k_1-k_0} B_{0,1}(x) = \frac{x-0}{0-0} B_{-1,1}(x) + \frac{2-x}{2-0} B_{0,1}(x)$$

Karena:

$$k_{-1}=k_0=0 \text{ maka } B_{-1,1}(x)=0 \text{ (sifat local support)}$$

$$B_{0,1}(x) = \begin{cases} 1 & , \quad 0 \leq x < 2 \\ 0 & , \quad \text{untuk } x \text{ yang lainnya} \end{cases}$$

sehingga:

$$B_{-1,2}(x) = \begin{cases} \frac{2-x}{2} & ; 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{0,2}(x) = \frac{x-k_0}{k_1-k_0}B_{0,1}(x) + \frac{k_2-x}{k_2-k_1}B_{1,1}(x) = \frac{x-0}{2-0}B_{0,1}(x) + \frac{4-x}{4-2}B_{1,1}(x)$$

Karena:

$$B_{0,1}(x) = \begin{cases} 1 & , 0 \leq x < 2 \\ 0 & , \text{ untuk } x \text{ yang lainnya} \end{cases}$$

$$B_{1,1}(x) = \begin{cases} 1 & , 2 \leq x < 4 \\ 0 & , \text{ untuk } x \text{ yang lainnya} \end{cases}$$

maka:

$$B_{0,2}(x) = \begin{cases} \frac{x}{2} & ; 0 \leq x < 2 \\ \frac{4-x}{2} & ; 2 \leq x < 4 \end{cases}$$

$$\text{Basis: } B_{1,2}(x) = \frac{x-k_1}{k_2-k_1}B_{1,1}(x) + \frac{k_3-x}{k_3-k_2}B_{2,1}(x) = \frac{x-2}{4-2}B_{1,1}(x) + \frac{4-x}{4-4}B_{2,1}(x)$$

Karena:

$$k_2 = k_3 = 4 \text{ maka } B_{2,1}(x) = 0 \text{ (sifat local support)}$$

$$B_{1,1}(x) = \begin{cases} 1 & , 2 \leq x < 4 \\ 0 & , \text{ untuk } x \text{ yang lainnya} \end{cases}$$

sehingga:

$$B_{1,2}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \frac{x-2}{2} & ; 2 \leq x < 4 \\ 0 & ; \text{ untuk } x \text{ yang lainnya} \end{cases}$$

Fungsi basis spline kuadratik ($m=3$) dengan $K=1$ knot pada $k_1=2$ dan nilai $x \in [0,4]$ memerlukan tambahan $2m=6$ knot yaitu $k_{-2}=k_{-1}=k_0=0$ dan $k_2=k_3=k_4=4$; sehingga knot menjadi $k_{-2}=k_{-1}=k_0=0, k_1=2, k_2=k_3=k_4=4$ dan matriks $B(x)$ berukuran $n \times 4$ sebagai berikut:

$$B(x) = \begin{bmatrix} B_{-2,3}(x) & B_{-1,3}(x) & B_{0,3}(x) & B_{1,3}(x) \end{bmatrix} = \begin{pmatrix} B_{-2,3}(x_1) & B_{-1,3}(x_1) & B_{0,3}(x_1) & B_{1,3}(x_1) \\ B_{-2,3}(x_2) & B_{-1,3}(x_2) & B_{0,3}(x_2) & B_{1,3}(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_{-2,3}(x_n) & B_{-1,3}(x_n) & B_{0,3}(x_n) & B_{1,3}(x_n) \end{pmatrix}$$

$$\text{Basis: } B_{-2,3}(x) = \frac{x-k_{-2}}{k_0-k_{-2}} B_{-2,2}(x) + \frac{k_1-x}{k_1-k_{-1}} B_{-1,2}(x) = \frac{x-0}{0-0} B_{-2,2}(x) + \frac{2-x}{2-0} B_{-1,2}(x)$$

Karena:

$$k_{-2} = k_0 = 0 \text{ maka } B_{-2,2}(x) = 0 \text{ (sifat local support)}$$

$$B_{-1,2}(x) = \begin{cases} \frac{2-x}{2} & ; 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{-2,3}(x) = \begin{cases} \left(\frac{2-x}{2} \right)^2 & ; 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{-1,3}(x) = \frac{x-k_{-1}}{k_1-k_{-1}} B_{-1,2}(x) + \frac{k_2-x}{k_2-k_0} B_{0,2}(x) = \frac{x-0}{2-0} B_{-1,2}(x) + \frac{4-x}{4-0} B_{0,2}(x)$$

Karena:

$$B_{-1,2}(x) = \begin{cases} \frac{2-x}{2} & ; 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$B_{0,2}(x) = \begin{cases} \frac{x}{2} & ; 0 \leq x < 2 \\ \frac{4-x}{2} & ; 2 \leq x < 4 \end{cases}$$

sehingga:

$$B_{-1,3}(x) = \begin{cases} \left(\frac{x}{2} \right) \left(\frac{2-x}{2} \right) + \left(\frac{4-x}{4} \right) \left(\frac{x}{2} \right) & ; 0 \leq x < 2 \\ \left(\frac{4-x}{4} \right) \left(\frac{4-x}{2} \right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{0,3}(x) = \frac{x-k_0}{k_2-k_0} B_{0,2}(x) + \frac{k_3-x}{k_3-k_1} B_{1,2}(x) = \frac{x-0}{4-0} B_{0,2}(x) + \frac{4-x}{4-2} B_{1,2}(x)$$

Karena:

$$B_{0,2}(x) = \begin{cases} \frac{x}{2} & ; 0 \leq x < 2 \\ \frac{4-x}{2} & ; 2 \leq x < 4 \end{cases}$$

$$B_{1,2}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \frac{x-2}{2} & ; 2 \leq x < 4 \\ 0 & ; \text{ untuk } x \text{ yang lainnya} \end{cases}$$

sehingga:

$$B_{0,3}(x) = \begin{cases} \left(\frac{x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{1,3}(x) = \frac{x-k_1}{k_3-k_1} B_{1,2}(x) + \frac{k_4-x}{k_4-k_2} B_{2,2}(x) = \frac{x-2}{4-2} B_{1,2}(x) + \frac{4-x}{4-4} B_{2,2}(x)$$

Karena:

$$k_2 = k_4 = 4 \text{ maka } B_{2,2}(x) = 0 \text{ (sifat local support)}$$

$$B_{1,2}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \frac{x-2}{2} & ; 2 \leq x < 4 \\ 0 & ; \text{ untuk } x \text{ yang lainnya} \end{cases}$$

sehingga:

$$B_{1,3}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \left(\frac{x-2}{2}\right)\left(\frac{x-2}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

Fungsi basis spline kubik ($m=4$) dengan $K=1$ knot pada $k_1=2$ dan nilai $x \in [0,4]$ memerlukan tambahan $2m=8$ knot sehingga knot menjadi $k_{-3}=k_{-2}=k_{-1}=k_0=0, k_1=2, k_2=k_3=k_4=k_5=4$ dan matriks $\mathbf{B}(x)$ berukuran $n \times 5$ sebagai berikut:

$$B(x) = \begin{bmatrix} B_{-3,4}(x) & B_{-2,4}(x) & B_{-1,4}(x) & B_{0,4}(x) & B_{1,4}(x) \end{bmatrix}$$

$$= \begin{pmatrix} B_{-3,4}(x_1) & B_{-2,4}(x_1) & B_{-1,4}(x_1) & B_{0,4}(x_1) & B_{1,4}(x_1) \\ B_{-3,4}(x_2) & B_{-2,4}(x_2) & B_{-1,4}(x_2) & B_{0,4}(x_2) & B_{1,4}(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{-3,4}(x_n) & B_{-2,4}(x_n) & B_{-1,4}(x_n) & B_{0,4}(x_n) & B_{1,4}(x_n) \end{pmatrix}$$

$$\text{Basis: } B_{-3,4}(x) = \frac{x-k_{-3}}{k_0-k_{-3}} B_{-3,3}(x) + \frac{k_1-x}{k_1-k_{-2}} B_{-2,3}(x) = \frac{x-0}{0-0} B_{-3,3}(x) + \frac{2-x}{2-0} B_{-2,3}(x)$$

Karena:

$$k_{-3} = k_0 = 0 \text{ maka } B_{-3,3}(x) = 0 \text{ (sifat local support)}$$

$$B_{-2,3}(x) = \begin{cases} \left(\frac{2-x}{2}\right)^2; & 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{-3,4}(x) = \begin{cases} \left(\frac{2-x}{2}\right)^3; & 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{-2,4}(x) = \frac{x-k_{-2}}{k_1-k_{-2}} B_{-2,3}(x) + \frac{k_2-x}{k_2-k_{-1}} B_{-1,3}(x) = \frac{x-0}{2-0} B_{-2,3}(x) + \frac{4-x}{4-0} B_{-1,3}(x)$$

Karena:

$$B_{-2,3}(x) = \begin{cases} \left(\frac{2-x}{2}\right)^2; & 0 \leq x < 2 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$B_{-1,3}(x) = \begin{cases} \left(\frac{x}{2}\right)\left(\frac{2-x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{2}\right); & 0 \leq x < 2 \\ \left(\frac{4-x}{4}\right)\left(\frac{4-x}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{-2,4}(x) = \begin{cases} \left(\frac{x}{2}\right)\left(\frac{2-x}{2}\right)^2 + \left(\frac{4-x}{4}\right)\left(\left(\frac{x}{2}\right)\left(\frac{2-x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{2}\right)\right); & 0 \leq x < 2 \\ \left(\frac{4-x}{4}\right)\left(\frac{4-x}{4}\right)\left(\frac{4-x}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{-1,4}(x) = \frac{x-k_{-1}}{k_2-k_{-1}} B_{-1,3}(x) + \frac{k_3-x}{k_3-k_0} B_{0,3}(x) = \frac{x-0}{4-0} B_{-1,3}(x) + \frac{4-x}{4-0} B_{0,3}(x)$$

Karena:

$$B_{-1,3}(x) = \begin{cases} \left(\frac{x}{2}\right)\left(\frac{2-x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{4-x}{4}\right)\left(\frac{4-x}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$B_{0,3}(x) = \begin{cases} \left(\frac{x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{-1,4}(x) = \begin{cases} \left(\frac{x}{4}\right)\left(\frac{x}{2}\right)\left(\frac{2-x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{x}{4}\right)\left(\frac{4-x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{4}\right)\left(\frac{x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{0,4}(x) = \frac{x-k_0}{k_3-k_0} B_{0,3}(x) + \frac{k_4-x}{k_4-k_1} B_{1,3}(x) = \frac{x-0}{4-0} B_{0,3}(x) + \frac{4-x}{4-2} B_{1,3}(x)$$

Karena:

$$B_{0,3}(x) = \begin{cases} \left(\frac{x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right) & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$B_{1,3}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \left(\frac{x-2}{2}\right)^2 & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{0,4}(x) = \begin{cases} \left(\frac{x}{4}\right)\left(\frac{x}{4}\right)\left(\frac{x}{2}\right) & ; 0 \leq x < 2 \\ \left(\frac{x}{4}\right)\left[\left(\frac{x}{4}\right)\left(\frac{4-x}{2}\right) + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right)\right] + \left(\frac{4-x}{2}\right)\left(\frac{x-2}{2}\right)^2 & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

$$\text{Basis: } B_{1,4}(x) = \frac{x-k_1}{k_4-k_1} B_{1,3}(x) + \frac{k_5-x}{k_5-k_2} B_{2,3}(x) = \frac{x-2}{4-2} B_{1,3}(x) + \frac{4-x}{4-4} B_{2,3}(x)$$

Karena:

$$k_2 = k_5 = 0 \text{ maka } B_{2,3}(x) = 0 \text{ (sifat local support)}$$

$$B_{1,3}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \left(\frac{x-2}{2}\right)^2 & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

sehingga:

$$B_{1,4}(x) = \begin{cases} 0 & ; 0 \leq x < 2 \\ \left(\frac{x-2}{2}\right)^3 & ; 2 \leq x < 4 \\ 0 & ; x \text{ lainnya} \end{cases}$$

C. REGRESI NONPARAMETRIK P-SPLINE

Dalam Ruppert (2003) disebutkan bahwa penggunaan knot yang relatif besar akan mempengaruhi kemulusan model, sehingga diperlukan suatu metode yang tetap mempertahankan banyak knot tetapi membatasi pengaruhnya.

Selain itu, penggunaan knot yang relatif besar akan menghasilkan penduga kurva regresi yang lebih bervariasi dalam memetakan data. Hasil yang sedikit fleksibel dapat diperoleh dengan menambahkan *penalty* pada derivatif ke dua dari taksiran kurva sehingga diperoleh fungsi *penalized least square (PLS)*:

$$PLS(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 + \lambda \int [f''(x)]^2 dx \quad (11)$$

dengan $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$ menyatakan rata-rata kuadrat sisaan, $\lambda \int [f''(x)]^2 dx$ menyatakan *roughness penalty*, yaitu ukuran pemulusan dalam memetakan data, dan $0 < \lambda < 1$ adalah parameter pemulus yang mengontrol keseimbangan antara kesesuaian terhadap data (*goodness of fit*) dan kemulusan kurva (*penalty*). Nilai λ yang besar (mendekati 1) akan memberikan bobot kemulusan kurva yang besar dan mempunyai variansi yang kecil.

Spline yang paling umum digunakan adalah potongan spline kubik ($m=4$). Jika k_1, k_2, \dots, k_K adalah kumpulan titik yang diurutkan sebagai knot,

maka spline kubik adalah fungsi kontinu $f(x_i)$ sedemikian sehingga $f(x_i)$ adalah polinomial kubik yang melalui (k_1, k_K) dan $f(x_i)$ mempunyai turunan pertama dan ke dua yang kontinu pada knot tersebut. Suatu spline linear yang melewati batas knot disebut *natural* spline. Fungsi $\hat{f}(x_i)$ yang meminimumkan $PLS(\lambda)$ dengan *penalty* $\lambda \int [f''(x)]^2 dx$ adalah spline kubik natural dengan knot pada titik-titik data dan penaksir $\hat{f}(x_i)$ disebut *smoothing* spline.

Fungsi basis yang digunakan dalam kajian ini adalah B-spline dengan matriks $B(x)$ berukuran $n \times (4 + K)$, yaitu:

$$B(x) = \begin{bmatrix} B_{-3,4}(x) & B_{-2,4}(x) & B_{-1,4}(x) & B_{0,4}(x) & B_{1,4}(x) \end{bmatrix}$$

$$= \begin{pmatrix} B_{-3,4}(x_1) & B_{-2,4}(x_1) & B_{-1,4}(x_1) & B_{0,4}(x_1) & B_{1,4}(x_1) \\ B_{-3,4}(x_2) & B_{-2,4}(x_2) & B_{-1,4}(x_2) & B_{0,4}(x_2) & B_{1,4}(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{-3,4}(x_n) & B_{-2,4}(x_n) & B_{-1,4}(x_n) & B_{0,4}(x_n) & B_{1,4}(x_n) \end{pmatrix}$$

dengan basis: $B_{j,4}(x) = \frac{x - k_j}{k_{j+3} - k_j} B_{j,3}(x) + \frac{k_{j+4} - x}{k_{j+4} - k_{j+1}} B_{j+1,3}(x)$ (12)

dan $B_{j,1}(x) = \begin{cases} 1 & , \quad x \in [k_j, k_{j+1}) \\ 0 & , \quad \text{untuk } x \text{ yang lainnya} \end{cases}$

Fungsi *PLS* pada Persamaan (11) dapat ditulis dalam bentuk matriks:

$$PLS(\lambda) = (Y - B\beta)'(Y - B\beta) + \lambda \beta' \Omega \beta \quad (13)$$

dengan $B_{ij} = B_j(x_i)$ dan $\Omega_{jk} = \int B_j''(x) B_k''(x) dx$.

Nilai β yang meminimumkan $PLS(\lambda)$ pada Persamaan (13) adalah:

$$\hat{\beta} = (B'B + \lambda\Omega)^{-1} B'Y \quad (14)$$

Smoothing spline $\hat{f}(x_i)$ adalah penghalus linear sehingga $\hat{f}(x_i)$ dapat dinyatakan

sebagai $\hat{f}(x_i) = \sum_{i=1}^n \ell(x_i) y_i$ untuk pembobot $\ell(x_i)$. Secara khusus, matriks

penghalus L dapat dinyatakan sebagai:

$$L = B(B'B + \lambda\Omega)^{-1} B' \quad (15)$$

dan vektor \hat{Y} adalah nilai taksiran yang dinyatakan sebagai:

$$\hat{Y} = LY \quad (16)$$

Jika dilakukan regresi linear sederhana dari \hat{Y} terhadap B maka matriks *hat* menjadi $L = B(B'B)^{-1}B'$ dan nilai taksiran \hat{Y} akan menginterpolasi data amatan. Pengaruh $\lambda\Omega$ pada Persamaan (15) adalah menyusutkan koefisien regresi menjadi suatu *subspace* yang menghasilkan dugaan penghalus. Derajat kebebasan yang efektif didefinisikan sebagai $\nu = \text{tr}(L) \leq n$ yang mempunyai nilai sama dengan banyaknya parameter pada persamaan regresi parametrik.

Parameter penghalus λ dipilih dengan cara meminimumkan nilai *generalized cross validation* (GCV) yang didefinisikan sebagai:

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}(x_i)}{1 - \frac{\nu}{n}} \right)^2 = \frac{\sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{n [1 - n^{-1} \text{tr}(L)]^2} = \frac{MSE(\lambda)}{(n^{-1} [\text{tr}(I - L)])^2} \quad (17)$$

Nilai penghalus dapat dibedakan menjadi *spar* dan λ , dimana $0 < spar \leq 1$ dan nilai λ diperoleh sebagai:

$$\lambda = r \cdot 256^{(3 \cdot spar - 1)} \quad \text{dengan} \quad r = \frac{\text{tr}(B'B)}{\text{tr}(\Omega)} \quad (18)$$

Nilai penghalus *spar* yang mendekati satu akan memberikan $\hat{f}(x_i)$ sebagai garis linear, sedangkan untuk nilai *spar* mendekati nol akan memberikan $\hat{f}(x_i)$ sebagai interpolasi. Nilai penghalus λ yang sangat besar akan menghasilkan bentuk kurva yang sangat halus, dan sebaliknya nilai λ yang kecil akan menghasilkan kurva yang kasar (Ruppert, 2003).

Dalam Wibowo (2011) disebutkan bahwa untuk memperoleh penaksir spline pada dasarnya terdapat dua pendekatan optimasi, yaitu penaksir spline yang diperoleh berdasarkan optimasi *penalized least square* (PLS) dan penaksir spline yang diperoleh berdasarkan optimasi *least square* (LS) dengan menggunakan fungsi keluarga yang memuat titik-titik knots. Apabila penaksir spline diperoleh berdasarkan optimasi PLS, maka persoalan utama dalam penaksiran adalah pemilihan parameter penghalus yang optimal. Jika penaksir spline diperoleh dengan optimasi LS, maka persoalan utama dalam penaksiran adalah pemilihan titik-titik knot yang optimal. Adapun pemilihan parameter penghalus dan pemilihan titik-titik knot yang optimal dapat menggunakan kriteria GVC, dimana GCV yang minimum akan memberikan parameter penghalus atau titik knot yang optimal (Eubank, 1999).

D. KRITERIA ASIMTOTIK

Kriteria asimtotik digunakan untuk mencari sifat-sifat penaksir dengan ukuran sampel n menuju tak hingga. Misalkan $T_n = l_n(x_1, x_2, \dots, x_n)$ merupakan barisan penaksir yang konsisten sederhana untuk θ jika untuk $\theta \in \Omega$ berlaku:

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| > \varepsilon) = 0, \forall \varepsilon > 0 \quad (19)$$

Dari pertidaksamaan *Chebysev* diperoleh:

$$P(|T_n - \theta| > \varepsilon) \leq E(T_n - \theta)^2 / \varepsilon^2 \quad (20)$$

Jika $T_n = l_n(x_1, x_2, \dots, x_n)$ merupakan penaksir tak bias untuk θ yaitu $E(T_n) = \theta$ maka:

$$P(|T_n - \theta| > \varepsilon) \leq E(T_n - \theta)^2 / \varepsilon^2 = \text{Var}(T_n) / \varepsilon^2 \quad (21)$$

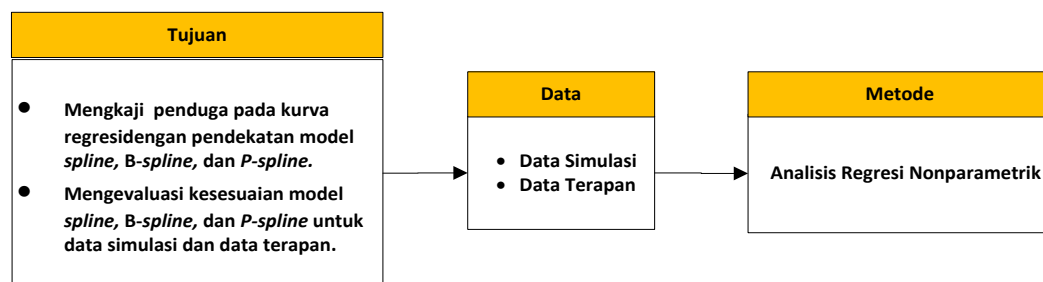
Berdasarkan Persamaan (18), $T_n = l_n(x_1, x_2, \dots, x_n)$ merupakan penaksir konsisten untuk parameter θ jika $\lim_{n \rightarrow \infty} \text{Var}(T_n) = 0$ dan $\lim_{n \rightarrow \infty} \text{bias}(T_n) = 0$ (Mood, Graybill dan Boes, 1974).

BAB III

METODE PENELITIAN

A. DESAIN PENELITIAN

Tujuan, data, metode, dan alat yang digunakan dalam penelitian ini disajikan dalam diagram berikut.



B. DATA

Guna memudahkan pemahaman tentang kajian metode yang ada, digunakan dua macam data, yaitu:

1. Data simulasi, berupa data bangkitan yang diperoleh dengan bantuan Software Program R versi 2.14.1 (R Development Core Team, 2011). Struktur data dengan pola linear, kuadratik, dan kubik diperoleh dengan membangkitkan sebanyak 100 pasang data (x, y) , dengan peubah bebas (x) berdistribusi uniform $(0 \leq x \leq 1)$ dan galat ε berdistribusi normal dengan mean 0 dan variansi 0,05.
2. Data terapan berupa data sekunder, yakni Indeks Prestasi (IP) pada semester satu dan lama studi mahasiswa PS Statistika FMIPA Universitas Terbuka yang lulus pada masa ujian 1989.2 sampai dengan 2011.2.

C. TAHAPAN ANALISIS

Adapun langkah-langkah yang akan dilakukan dalam penelitian ini adalah:

1. Mengkaji penduga pada kurva regresi dengan pendekatan model spline dengan langkah-langkah:
 - a. Membangkitkan data simulasi sebagai variabel prediktor (x) sebanyak 100 dengan berdistribusi uniform ($0 \leq x \leq 1$) dan galat $\varepsilon \sim NIID(0; 0,05)$.
 - b. Mendeskripsikan pola hubungan antara variabel respon (y) terhadap variabel prediktor (x).
 - c. Menentukan jumlah titik knot dan letak titik knot.
 - d. Menghitung nilai taksiran parameter.
 - e. Menghitung nilai MSE .
 - f. Menentukan titik knot dan orde knot yang menghasilkan nilai MSE terkecil untuk setiap model.
 - g. Membuat model regresi nonparametrik spline berdasarkan titik knot dan orde knot yang menghasilkan nilai MSE terkecil.
2. Mengkaji penduga pada kurva regresi dengan pendekatan model B-spline dengan langkah-langkah:
 - a. Membuat fungsi basis B-spline yakni matriks $B = \{B_{j,m} ; j = 1, 2, \dots, K + m\}$ berdasarkan data simulasi berpasangan yang telah dibangkitkan.
 - b. Menentukan jumlah titik knot dan letak titik knot.
 - c. Menghitung nilai taksiran parameter.
 - d. Menghitung nilai MSE .
 - e. Menentukan titik knot dan orde knot yang menghasilkan nilai MSE terkecil untuk setiap model.
 - f. Membuat model regresi nonparametrik spline berdasarkan titik knot dan orde knot yang menghasilkan nilai MSE terkecil.

3. Mengkaji penduga pada kurva regresi dengan pendekatan model P-spline dengan langkah-langkah:
 - a. Membuat fungsi basis yakni matriks $B = \{B_{j,m} ; j = 1, 2, \dots, K + m\}$ secara iteratif berdasarkan data simulasi berpasangan yang telah dibangkitkan.
 - b. Menghitung $r = \frac{tr(B'B)}{tr(\Omega)}$ dengan $\Omega_{jk} = \int B_j''(x) B_k''(x) dx$.
 - c. Menentukan nilai penghalus $\lambda = r \cdot 256^{(3spar-1)}$ dengan nilai $0 < spar \leq 1$, dan $spar_0 = 0$.
 - d. Menentukan matriks penghalus $L = B(B'B + \lambda\Omega)^{-1} B'$ dan $v = tr(L)$.
 - e. Menghitung taksiran parameter $\hat{\beta} = (B'B + \lambda\Omega)^{-1} B'Y$.
 - f. Menghitung nilai GCV.
 - g. Menentukan nilai GCV terkecil untuk setiap model.
 - h. Membuat model regresi nonparametrik P-spline berdasarkan titik knot dan orde knot yang menghasilkan nilai GCV terkecil.
 - i. Membandingkan nilai MSE dan GCV dari model spline, B-spline, dan P-spline.
4. Mendapatkan model regresi nonparametrik menggunakan pendekatan spline, B-spline, dan P-spline dari data real berupa Indeks Prestasi (IP) pada semester satu dan IPK kelulusan mahasiswa PS Statistika yang lulus pada masa ujian 1989.2 sampai dengan 2011.2 dengan langkah-langkah :
 - a. Menghitung nilai taksiran dari model spline, B-spline, dan P-spline.
 - b. Melakukan pengujian signifikansi koefisien regresi secara parsial dan simultan.
 - c. Menyimpulkan dan menginterpretasikan hasil taksiran.

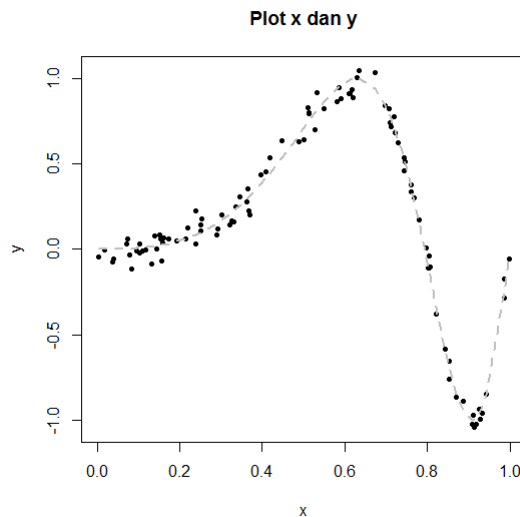
BAB IV

HASIL DAN PEMBAHASAN

A. HASIL SIMULASI

1. *Spline dengan Basis Truncated*

Berdasarkan data simulasi peubah bebas (x) berdistribusi uniform ($0 \leq x \leq 1$) dan galat ε berdistribusi normal dengan mean 0 dan variansi 0,05 berukuran 100 diperoleh peubah tak bebas (y) dengan fungsi $y = \sin(2\pi x^3) + \varepsilon$. Plot antara peubah x dan y dapat dilihat pada Gambar 1 dan garis putus-putus menunjukkan persamaan garis sesungguhnya.



Gambar 1. Plot x dan y

Selanjutnya, dengan memperhatikan pola hubungan variabel respon y terhadap variabel prediktor x tampak bahwa ada perubahan pola dari peubah bebas pada interval tertentu. Jika digunakan titik $k_1 = 0,3$ sebagai knot, maka taksiran parameter fungsi spline berorde dua (linear) dengan basis *truncated* adalah: $\hat{f}(x) = -0,285 + 3,109x - 4,408(x - 0,3)$ dengan nilai $MSE = 0,204$. Selain taksiran parameter yang diperoleh sangat signifikan, secara simultan model ini signifikan pada $\alpha = 5\%$ seperti ditunjukkan pada Tabel 1 berikut.

Tabel 1. Taksiran Fungsi Spline Linear dengan knot $k_1 = 0,3$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
<i>Intercept</i>	- 0.285	0.139	0.0426 *	0.272 8	18.19 0	1.96E-07	0.204
<i>x</i>	3.109	0.624	2.76e-06 ***				
<i>x-0,3</i>	- 4.408	0.780	1.59e-07 ***				

Jika digunakan titik $k_1 = 0,3$ dan $k_2 = 0,6$ sebagai knot, maka taksiran fungsi spline linear adalah: $\hat{f}(x) = -0,043 + 0,579x + 2,502(x - 0,3) - 8,058(x - 0,6)$ dengan nilai $MSE = 0,053$. Secara parsial taksiran untuk parameter intercept dan peubah x tidak signifikan pada $\alpha = 5\%$, tetapi taksiran parameter secara simultan sangat signifikan dan nilai MSE yang diperoleh lebih kecil dibanding pada model sebelumnya seperti ditunjukkan pada Tabel 2.

Tabel 2. Taksiran Fungsi Spline Linear dengan knot $k_1 = 0,3$ dan $k_2 = 0,6$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
<i>Intercept</i>	-0.04308	0.07221	0.552	0.812 4	138. 6	< 2.2e-16	0.053
<i>x</i>	0.57893	0.35315	0.104				
<i>x-0,3</i>	2.50156	0.57572	3.46e-05 ***				
<i>x-0,6</i>	-8.05828	0.48492	< 2e-16 ***				

Selanjutnya jika digunakan tiga knot yaitu $k_1 = 0,3$, $k_2 = 0,6$, dan $k_3 = 0,9$, maka akan diperoleh taksiran fungsi spline linear:

$$\hat{f}(x) = -0,026 + 0,396x + 3,361(x - 0,3) - 10,704(x - 0,6) + 12,699(x - 0,9).$$

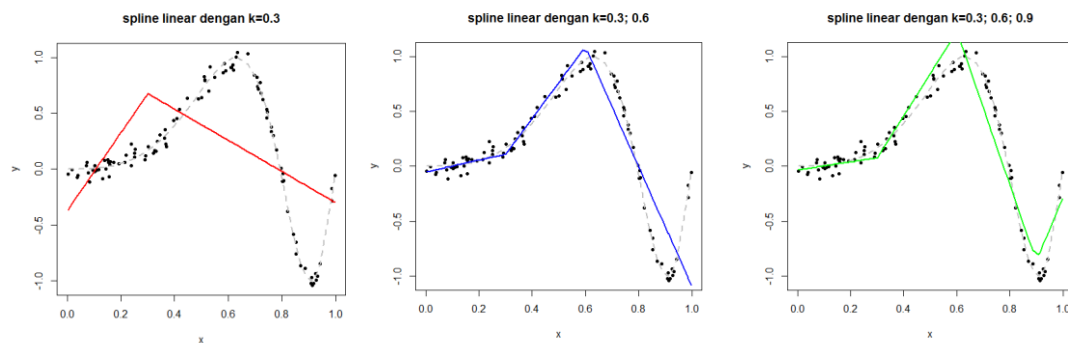
Meskipun secara parsial, taksiran parameternya untuk intercept dan peubah x tidak signifikan, namun secara simultan, model ini signifikan dengan nilai $MSE = 0,023$. Secara lengkap, taksiran fungsi dan kurva taksiran fungsi spline linear untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Tabel 3 dan Gambar 2 berikut.

Tabel 3. Taksiran Fungsi Spline Linear dengan Knot $k_1 = 0,3$, $k_2 = 0,6$, dan $k_3 = 0,9$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
<i>Intercept</i>	-0.02562	0.04711	0.5878	0.921 1	277. 2	< 2.2e-16	0.023
<i>x</i>	0.39604	0.23081	0.0894.				
<i>x-0,3</i>	3.36096	0.38282	6.7e-14 ***				

$x=0,6$	10.70372	0.39174	$< 2e-16$ ***
$x=0,9$	12.69907	1.11028	$< 2e-16$ ***

Pada Gambar 2 tampak bahwa taksiran fungsi spline linear dengan tiga knot lebih sesuai atau mendekati garis sesungguhnya (garis putus-putus) dibanding taksiran fungsi spline linear dengan satu atau dua knot.



Gambar 2. Kurva Spline Linear

Taksiran fungsi spline berorde tiga (kuadratik) dengan menggunakan satu knot $k_1 = 0,3$ adalah $\hat{f}(x) = 0,120 - 3,444x + 12,791x^2 - 21,557(x - 0,3)^2$ dengan $MSE = 0,078$. Pada Tabel 4 dapat dilihat bahwa secara parsial, taksiran parameter untuk semua peubah (kecuali intercept) signifikan pada $\alpha = 5\%$ dan secara simultan model ini sangat signifikan. Jika dibandingkan dengan taksiran fungsi spline linear dengan dua knot atau tiga knot, taksiran fungsi spline kuadratik dengan satu knot mempunyai nilai MSE yang lebih besar. Hal ini memberikan indikasi bahwa taksiran fungsi spline kuadratik dengan satu knot masih belum optimal, sehingga perlu dicoba untuk menentukan taksiran fungsi spline kuadratik dengan dua knot dan tiga knot.

Tabel 4. Taksiran Fungsi Spline Kuadratik dengan Knot $k_1 = 0,3$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
				0.695			
<i>Intercept</i>	0.1195	0.1302	0.36089	9	73.24	$< 2.2e-16$	0.078
<i>x</i>	-3.444	1.2107	0.00544 **				
<i>x^2</i>	12.7914	2.4788	1.33e-06 ***				
<i>(x-0,3)^2</i>	21.5571	2.9589	7.25e-11 ***				

Taksiran fungsi spline kuadratik dengan dua knot $k_1=0,3$ dan $k_2=0,6$ adalah $\hat{f}(x)=0,103-3,058x+11,672x^2-19,529(x-0,3)^2-2,287(x-0,6)^2$ dengan $MSE=0,079$. Tidak jauh berbeda dengan taksiran fungsi spline kuadratik menggunakan satu knot, pada Tabel 5 dapat dilihat bahwa secara parsial, taksiran parameter fungsi spline kuadratik dengan dua knot (kecuali intercept) signifikan pada $\alpha=5\%$ dan secara simultan model ini sangat signifikan.

Tabel 5. Taksiran Fungsi Spline Kuadratik dengan Knot $k_1=0,3$ dan $k_2=0,6$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
				0.697	54.6		
<i>intercept</i>	0.103	0.1332	0.441353	2	8	< 2.2e-16	0.079
<i>x</i>	-3.0577	1.3627	0.027159 *				
<i>x^2</i>	11.6722	3.064	0.000247 ***				
	-						
<i>(x-0,3)^2</i>	19.5294	4.5523	4.3e-05 ***				
<i>(x-0,6)^2</i>	-2.2868	3.6576	0.533324				

Selanjutnya pada Tabel 6 dapat dilihat bahwa taksiran fungsi spline kuadratik dengan tiga knot $k_1=0,3$, $k_2=0,6$ dan $k_3=0,9$ adalah:

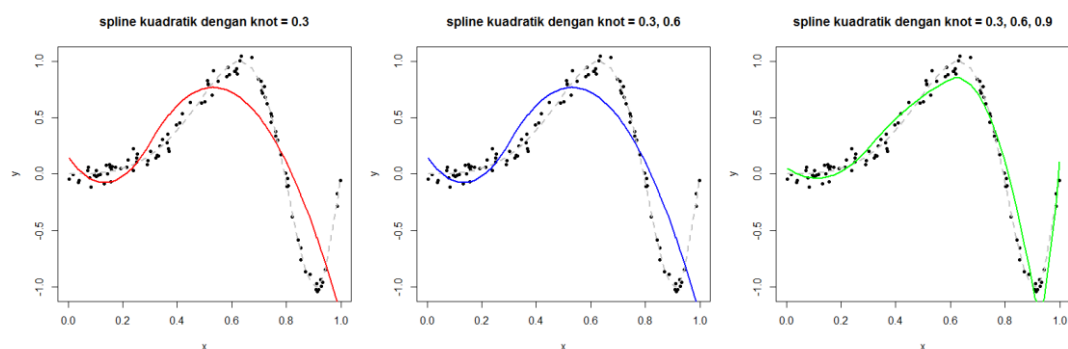
$\hat{f}(x)=0,039-1,477x+6,767x^2-8,047(x-0,3)^2-26,433(x-0,6)^2+305,744(x-0,9)^2$ dengan $MSE=0,015$. Tidak jauh berbeda dengan taksiran fungsi spline kuadratik sebelumnya, secara parsial, taksiran parameter fungsi spline kuadratik dengan tiga knot (kecuali intercept) signifikan pada $\alpha=5\%$ dan secara simultan model sangat signifikan. Namun, fungsi spline kuadratik dengan tiga knot mempunyai nilai MSE yang paling kecil diantara fungsi spline sebelumnya. Hal ini memberikan indikasi bahwa fungsi spline kuadratik dengan tiga knot mempunyai performa yang paling bagus diantara fungsi spline sebelumnya.

Tabel 6. Taksiran Fungsi Spline Kuadratik dengan Knot $k_1=0,3$, $k_2=0,6$, dan $k_3=0,9$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
				0.944	317.		
<i>intercept</i>	0.03871	0.05765	0.503547	1	2	< 2.2e-16	0.015
<i>x</i>	-1.47693	0.59391	0.014654 *				
<i>x^2</i>	6.76672	1.3457	2.36e-06 ***				
<i>(x-0,3)^2</i>	-8.04685	2.04626	0.000161 ***				
	-						
<i>(x-0,6)^2</i>	26.43291	1.97571	< 2e-16 ***				

$$(x-0,9)^2 \quad 305.7439 \quad 15.0119 \quad 9 \quad < 2e-16 ***$$

Kurva taksiran fungsi spline kuadratik untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 3 berikut. Pada Gambar 3 tampak bahwa taksiran fungsi spline kuadratik dengan tiga knot lebih mendekati garis sesungguhnya (garis putus-putus) dibanding taksiran fungsi spline kuadratik dengan satu atau dua knot.



Gambar 3. Kurva Spline Kuadratik

Taksiran fungsi spline orde empat (kubik) dengan menggunakan satu knot $k_1 = 0,3$ adalah $\hat{f}(x) = 0,114 - 5,415x + 31,916x^2 - 39,842x^3 + 34,923(x - 0,3)^3$ dan $MSE = 0,085$ seperti ditunjukkan pada Tabel 7. Secara parsial, taksiran parameter untuk semua peubah (kecuali intercept) signifikan pada $\alpha = 5\%$ dan secara simultan model ini sangat signifikan. Jika dibandingkan dengan taksiran fungsi spline kuadratik, taksiran fungsi spline kubik dengan satu knot mempunyai nilai MSE yang lebih besar. Hal ini memberikan indikasi bahwa performa fungsi spline kubik dengan satu knot masih belum optimal, sehingga perlu dicoba untuk menentukan taksiran fungsi spline kubik dengan dua knot dan tiga knot.

Tabel 7. Taksiran Fungsi Spline Kubik dengan Knot $k_1 = 0,3$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
				0.671	48.5		
<i>Intercept</i>	0.1137	0.1562	0.46842	5	5	< 2.2e-16	0.085
<i>x</i>	-5.4153	2.5758	0.03816 *				
<i>x^2</i>	31.9162	11.7032	0.00761 **				
	-						
<i>x^3</i>	39.8423	15.2569	0.01048 *				
<i>(x-0,3)^3</i>	34.9225	17.6454	0.05070 .				

Taksiran fungsi spline kubik dengan menggunakan dua knot $k_1 = 0,3$ dan $k_2 = 0,6$ dapat dilihat pada Tabel 8 yaitu:

$\hat{f}(x) = -0,161 + 6,402x - 41,347x^2 + 74,351x^3 - 143,566(x-0,3)^3 + 149,927(x-0,6)^3$
dan $MSE = 0,025$. Secara parsial, taksiran parameter untuk semua peubah (kecuali intercept) signifikan pada $\alpha = 5\%$ dan secara simultan model ini sangat signifikan. Jika dibandingkan dengan taksiran fungsi spline kuadratik dengan tiga knot, taksiran fungsi spline kubik dengan dua knot mempunyai nilai MSE yang lebih besar. Hal ini memberikan indikasi bahwa performa fungsi spline kubik dengan dua knot masih belum optimal.

Tabel 8. Taksiran Fungsi Spline Kubik dengan Knot $k_1 = 0,3$ dan $k_2 = 0,6$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i> 0.903	<i>F</i> 175.	<i>p-value</i>	<i>MSE</i>
<i>Intercept</i>	-0.16112	0.08724	0.06790 .	1	2	< 2.2e-16	0.025
<i>x</i>	6.40187	1.61228	0.00014 ***				
<i>x^2</i>	-41.34671	8.04502	1.49e-06 ***				
<i>x^3</i>	74.35132	11.28878	2.57e-09 ***				
	-						
<i>(x-0,3)^3</i>	143.56604	15.31747	3.97e-15 ***				
<i>(x-0,6)^3</i>	149.92651	10.00255	< 2e-16 ***				

Taksiran fungsi spline kubik dengan menggunakan tiga knot $k_1 = 0,3$, $k_2 = 0,6$ dan $k_3 = 0,9$ dapat dilihat pada Tabel 9 yaitu:

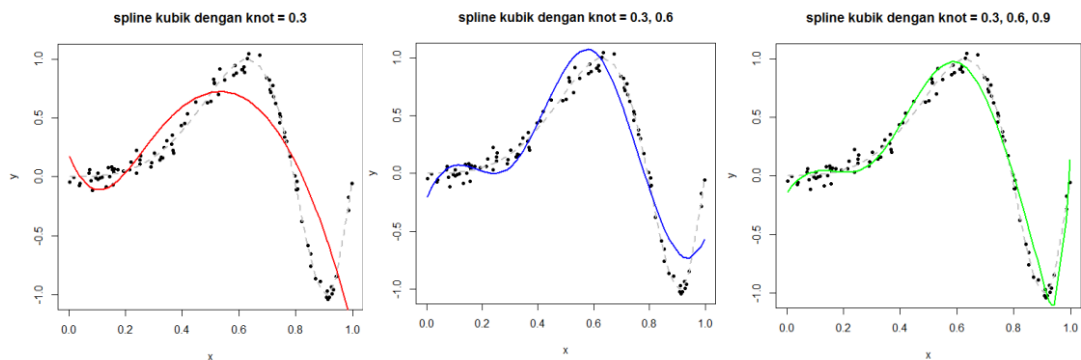
$\hat{f}(x) = -0,145 + 3,617x - 22,506x^2 + 44,204x^3 - 92,634(x-0,3)^3 + 75,30(x-0,6)^3 + 2013,961(x-0,9)^3$
dan $MSE = 0,013$. Secara parsial, taksiran parameter untuk semua peubah signifikan pada $\alpha = 5\%$ dan secara simultan model ini sangat signifikan. Jika dibandingkan dengan taksiran fungsi spline kuadratik dengan tiga knot, taksiran fungsi spline kubik dengan dua knot mempunyai nilai MSE yang hampir sama. Hal ini memberikan indikasi bahwa performa fungsi spline kubik dengan tiga knot sudah optimal.

Tabel 9. Taksiran Fungsi Spline Kubik dengan Knot $k_1 = 0,3, k_2 = 0,6$ dan $k_3 = 0,9$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i> 0.959	<i>F</i> 364.	<i>p-value</i>	<i>MSE</i>
<i>Intercept</i>	-0.14479	0.07276	0.049538 *	2	7	< 2.2e-16	0.013
<i>x</i>	3.61742	1.25791	0.004997 **				
<i>x^2</i>	-22.50615	6.07889	0.000362 ***				

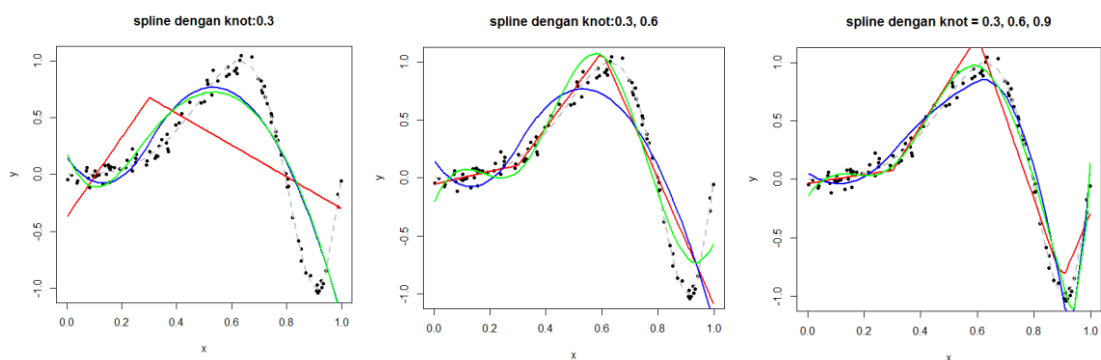
x^3	44.20447	8.49112	1.15e-06 ***
		11.6686	
$(x-0,3)^3$	-92.63352	4	4.55e-12 ***
$(x-0,6)^3$	75.32999	9.04392	6.92e-13 ***
	2013.9612	161.179	
$(x-0,9)^3$	2	1	< 2e-16 ***

Kurva taksiran fungsi spline kubik untuk satu knot $k_1=0,3$, untuk dua knot $k_1=0,3; k_2=0,6$, dan untuk tiga knot $k_1=0,3; k_2=0,6; k_3=0,9$ dapat dilihat pada Gambar 4 berikut. Pada Gambar 4 tampak bahwa taksiran fungsi spline kubik dengan tiga knot yang paling mendekati garis sesungguhnya (garis putus-putus) dibanding taksiran fungsi spline kubik dengan satu atau dua knot.



Gambar 4. Kurva Spline Kubik

Secara gabungan, kurva taksiran fungsi spline linear, kuadratik, dan kubik untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 5.



Gambar 5. Kurva Spline Linear, Kuadratik, dan Kubik

Jika dilihat dari Gambar 5, kurva taksiran fungsi spline linear, kuadratik, dan kubik untuk tiga knot $k_1=0,3; k_2=0,6$ dan $k_3=0,9$ lebih mendekati kurva sesungguhnya. Lebih khusus lagi, kurva taksiran fungsi spline kuadratik dan

kubik untuk tiga knot $k_1 = 0,3$; $k_2 = 0,6$ dan $k_3 = 0,9$ mempunyai performa yang sama baiknya.

2. B-spline

Kurva spline orde dua (linear) dengan fungsi basis B-spline dan satu knot yaitu $k_1 = 0,3$ mempunyai tiga koefisien basis yang harus ditaksir. Jika $B_{-1,2}$, $B_{0,2}$, $B_{1,2}$ merupakan basis spline linear dengan satu knot, maka taksiran kurva regresi adalah: $\hat{f}(x) = -0,365 B_{-1,2}(x) + 1,042 B_{0,2}(x) + 0,067 B_{1,2}(x)$ dengan $MSE = 0,217$. Secara parsial taksiran untuk semua koefisien signifikan pada $\alpha = 5\%$ kecuali koefisien basis $B_{1,2}$, tetapi secara simultan sangat signifikan seperti ditunjukkan pada Tabel 10.

Tabel 10. Taksiran Fungsi B-spline Linear dengan Knot $k_1 = 0,3$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
	-			0.268	17.8		
$B_{-1,2}$	0.36519	0.15351	0.0193 *	7	2	2.56E-07	0.217
$B_{0,2}$	1.04195	0.20947	2.83e-06 ***				
$B_{1,2}$	0.06662	0.17227	0.6998				

Pada Tabel 11 dapat dilihat taksiran kurva B-spline linear dengan $k_1 = 0,3$ dan $k_2 = 0,6$ sebagai knot serta basis $B_{-1,2}$, $B_{0,2}$, $B_{1,2}$, $B_{2,2}$ yaitu:

$\hat{f}(x) = -0,052 B_{-1,2}(x) + 0,161 B_{0,2}(x) + 1,142 B_{1,2}(x) - 1,032 B_{2,2}(x)$ dengan nilai $MSE = 0,055$. Secara parsial, taksiran untuk koefisien basis $B_{-1,2}$ dan $B_{0,2}$ tidak signifikan pada $\alpha = 5\%$, tetapi secara simultan sangat signifikan dan nilai MSE yang diperoleh lebih kecil dibanding kurva spline linear dengan satu knot.

Tabel 11. Taksiran Fungsi B-spline Linear dengan Knot $k_1 = 0,3$ dan $k_2 = 0,6$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
	-			0.816	142.		
$B_{-1,2}$	0.05178	0.07939	0.516	9	8	< 2.2e-16	0.055
$B_{0,2}$	0.16101	0.11747	0.174				
$B_{1,2}$	1.14178	0.09297	< 2e-16 ***				
	-						
$B_{2,2}$	1.03161	0.10818	1.47e-15 ***				

Kurva B-spline linear dengan tiga knot $k_1=0,3$, $k_2=0,6$, dan $k_3=0,9$ mempunyai lima basis yaitu $B_{-1,2}$, $B_{0,2}$, $B_{1,2}$, $B_{2,2}$, $B_{3,2}$ dan taksirannya dapat dilihat pada Tabel 12, yaitu:

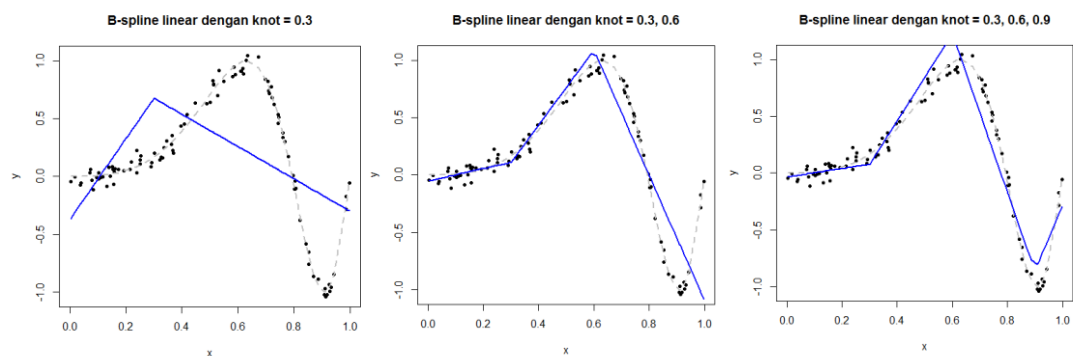
$$\hat{f}(x) = -0,034 B_{-1,2}(x) + 0,110 B_{0,2}(x) + 1,263 B_{1,2}(x) - 0,830 B_{2,2}(x) - 0,262 B_{3,2}(x)$$

dengan $MSE=0,030$. Secara parsial, taksiran untuk koefisien basis $B_{-1,2}$ dan $B_{0,2}$ tidak signifikan pada $\alpha=5\%$, tetapi secara simultan sangat signifikan dan nilai MSE yang diperoleh lebih kecil dibanding kurva B-spline linear dengan satu dan dua knot, sehingga performa kurva B-spline linear dengan tiga knot cukup baik.

Tabel 12. Taksiran Fungsi B-spline Linear dengan Knot $k_1=0,3$, $k_2=0,6$, dan $k_3=0,9$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
$B_{-1,2}$	-0.03374	0.05853	0.566	0.901	217.6	< 2.2e-16	0.030
$B_{0,2}$	0.1103	0.08673	0.207				
$B_{1,2}$	1.26273	0.06979	<2e-16 ***				
$B_{2,2}$	-0.82997	0.07837	<2e-16 ***				
$B_{3,2}$	-0.26189	0.11659	0.027 *				

Kurva taksiran B-spline linear untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 6 berikut. Tampak bahwa taksiran fungsi B-spline linear dengan tiga knot lebih mendekati kurva sesungguhnya.



Gambar 6. Kurva B-spline Linear

Kurva B-spline orde tiga (kuadratik) dengan satu knot yaitu $k_1=0,3$ mempunyai empat koefisien basis yang harus ditaksir. Jika $B_{-2,3}, B_{-1,3}, B_{0,3}, B_{1,3}$ merupakan basis spline kuadratik dengan satu knot, maka taksiran kurva B-spline dapat dilihat pada Tabel 13, yaitu: $\hat{f}(x)=0,140 B_{-2,3}(x)-0,495 B_{-1,3}(x)+1,621 B_{0,3}(x)-1,376 B_{1,3}(x)$ dan $MSE=0,083$. Secara simultan, taksiran koefisien basis spline signifikan pada $\alpha=5\%$ tetapi secara parsial taksiran koefisien basis $B_{-2,3}$ tidak signifikan.

Tabel 13. Taksiran Fungsi B-spline Kuadratik dengan Knot $k_1=0,3$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
				0.724			
$B_{-2,3}$	0.1402	0.138	0.3121	1	84	< 2.2e-16	0.083
	-						
$B_{-1,3}$	0.4951	0.2052	0.0177 *				
$B_{0,3}$	1.6206	0.1548	< 2e-16 ***				
	-						
$B_{1,3}$	1.3759	0.1808	1.88e-11 ***				

Kurva B-spline kuadratik dengan dua knot $k_1=0,3$ dan $k_2=0,6$ mempunyai lima koefisien basis yang harus ditaksir yaitu koefisien untuk basis $B_{-2,3}, B_{-1,3}, B_{0,3}, B_{1,3}, B_{2,3}$. Pada Tabel 14 dapat dilihat taksiran kurvanya adalah: $\hat{f}(x)=0,140 B_{-2,3}(x)-0,495 B_{-1,3}(x)+0,775 B_{0,3}(x)+0,333 B_{1,3}(x)-1,376 B_{2,3}(x)$ dan nilai $MSE=0,084$ yang tidak jauh berbeda dengan kurva B-spline kuadratik satu knot $k_1=0,3$. Secara simultan, taksiran koefisien basis spline signifikan pada $\alpha=5\%$ tetapi secara parsial taksiran koefisien basis $B_{-2,3}$ dan $B_{1,3}$ tidak signifikan.

Tabel 14. Taksiran Fungsi B-spline Kuadratik dengan Knot $k_1=0,3$ dan $k_2=0,6$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
				0.724	62.3		
$B_{-2,3}$	0.1402	0.1455	0.3378	1	5	< 2.2e-16	0.084
	-						
$B_{-1,3}$	0.4951	0.2383	0.0404 *				
$B_{0,3}$	0.775	0.1551	2.65e-06 ***				
$B_{1,3}$	0.3329	0.2013	0.1015				
	-						
$B_{2,3}$	1.3759	0.1841	3.82e-11 ***				

Kurva B-spline kuadratik dengan tiga knot yaitu $k_1 = 0,3$, $k_2 = 0,6$, dan $k_3 = 0,9$ mempunyai enam basis yaitu: $B_{-2,3}$, $B_{-1,3}$, $B_{0,3}$, $B_{1,3}$, $B_{2,3}$, $B_{3,3}$. Tabel 15 menunjukkan taksiran kurva B-spline adalah:

$$\hat{f}(x) = 0,051 B_{-2,3}(x) - 0,243 B_{-1,3}(x) + 0,610 B_{0,3}(x) + 0,965 B_{1,3}(x) - 1,696 B_{2,3}(x) + 0,057 B_{3,3}(x)$$

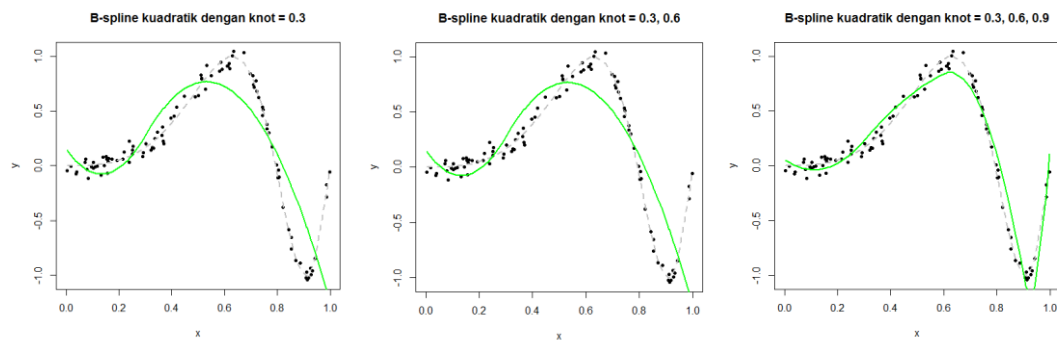
dan $MSE = 0,015$. Secara simultan, taksiran koefisien basis spline signifikan pada $\alpha = 5\%$ tetapi secara parsial taksiran koefisien basis $B_{-2,3}$ dan $B_{3,3}$ tidak signifikan. Nilai MSE yang diperoleh lebih kecil dibanding kurva B-spline kuadratik dengan satu knot dan dua knot, sehingga performa kurva B-spline kuadratik dengan tiga knot lebih baik dibanding kurva B-spline kuadratik dengan satu dan dua knot.

Tabel 15. Taksiran Fungsi B-spline Kuadratik dengan Knot

$$k_1 = 0,3, k_2 = 0,6, \text{ dan } k_3 = 0,9$$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
$B_{-2,3}$	0.05111	0.06188	0.411	0.950	364.		
				9	1	< 2.2e-16	0.015
$B_{-1,3}$	-0.24306	0.1018	0.019 *				
$B_{0,3}$	0.60972	0.06625	9.11e-15 ***				
$B_{1,3}$	0.96448	0.08212	< 2e-16 ***				
$B_{2,3}$	-1.69603	0.07849	< 2e-16 ***				
$B_{3,3}$	0.05657	0.10404	0.588				

Kurva taksiran B-spline kuadratik untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 7. Tampak bahwa taksiran fungsi B-spline kuadratik dengan tiga knot lebih mendekati kurva sesungguhnya (garis putus-putus).



Gambar 7. Kurva B-spline Kuadratik

Kurva B-spline orde empat (kubik) dengan satu knot yaitu $k_1=0,3$ mempunyai lima koefisien basis yang harus ditaksir.

Tabel 16. Taksiran Fungsi B-spline Kubik dengan Knot $k_1=0,3$

<i>Basis</i>	<i>Coef</i>	<i>SD(coef)</i>	<i>Pr(> t)</i>	<i>R2</i>	<i>F</i>	<i>p-value</i>	<i>MSE</i>
$B_{-3,4}$	0.1721	0.182	0.346541	0.700	55.5	< 2.2e-16	0.091
				6	8		
$B_{-2,4}$	-0.5867	0.2895	0.045519 *				
$B_{-1,4}$	0.9469	0.2515	0.000288 ***				
$B_{0,4}$	0.7241	0.3014	0.018222 *				
$B_{1,4}$	-1.4303	0.213	1.38e-09 ***				

Jika $B_{-3,4}, B_{-2,4}, B_{-1,4}, B_{0,4}, B_{1,4}$ dan merupakan basis spline kubik dengan satu knot, maka taksiran kurva B-spline dapat dilihat pada Tabel 16, yaitu:

$$\hat{f}(x) = 0,172 B_{-3,4}(x) - 0,587 B_{-2,4}(x) + 0,947 B_{-1,4}(x) + 0,724 B_{0,4}(x) - 1,430 B_{1,4}(x)$$

dan $MSE = 0,091$. Secara simultan, taksiran koefisien basis spline signifikan pada $\alpha = 5\%$ tetapi secara parsial taksiran koefisien basis $B_{-3,4}$ tidak signifikan.

Kurva B-spline kubik dengan dua knot yaitu $k_1=0,3$ dan $k_2=0,6$ mempunyai enam basis $B_{-3,4}, B_{-2,4}, B_{-1,4}, B_{0,4}, B_{1,4}, B_{2,4}$ dan taksiran kurvanya dapat dilihat pada Tabel 17, yaitu: $\hat{f}(x) = -0,201 B_{-3,4}(x) + 0,561 B_{-2,4}(x) - 0,482 B_{-1,4}(x) + 2,683 B_{0,4}(x) - 1,135 B_{1,4}(x) - 0,374 B_{2,4}(x)$ dan $MSE = 0,033$. Secara simultan, taksiran koefisien basis spline signifikan pada $\alpha = 5\%$ tetapi secara parsial taksiran koefisien basis $B_{-3,4}$ tidak signifikan.

Tabel 17. Taksiran Fungsi B-spline Kubik dengan Knot $k_1 = 0,3$ dan $k_2 = 0,6$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
$B_{-3,4}$	-0.201	0.1143	0.081752 .	0.8908	153.4	< 2.2e-16	0.033
$B_{-2,4}$	0.5611	0.1973	0.005480 **				
$B_{-1,4}$	-0.4815	0.1336	0.000503 ***				
$B_{0,4}$	2.6832	0.1855	< 2e-16 ***				
$B_{1,4}$	1.1353	0.1519	4e-11 ***				
$B_{2,4}$	0.3735	0.1535	0.016823 *				

Kurva B-spline kubik dengan tiga knot yaitu $k_1 = 0,3$, $k_2 = 0,6$, dan $k_3 = 0,9$ mempunyai tujuh basis yaitu: $B_{-3,4}$, $B_{-2,4}$, $B_{-1,4}$, $B_{0,4}$, $B_{1,4}$, $B_{2,4}$, $B_{3,4}$. Tabel 18 menunjukkan taksiran kurva B-spline adalah:

$$\hat{f}(x) = -0,137 B_{-3,4}(x) + 0,350 B_{-2,4}(x) - 0,268 B_{-1,4}(x) + 1,897 B_{0,4}(x) - 0,093 B_{1,4}(x) - 1,339 B_{2,4}(x) + 0,274 B_{3,4}(x)$$

dan $MSE = 0,013$. Secara simultan, taksiran koefisien basis signifikan pada $\alpha = 5\%$ tetapi secara parsial taksiran koefisien basis $B_{-3,4}$ dan $B_{1,4}$ tidak signifikan. Nilai MSE yang diperoleh lebih kecil dibanding kurva B-spline kubik dengan satu knot dan dua knot, sehingga performa kurva B-spline kubik dengan tiga knot lebih baik dibanding kurva B-spline kubik dengan satu knot dan dua knot.

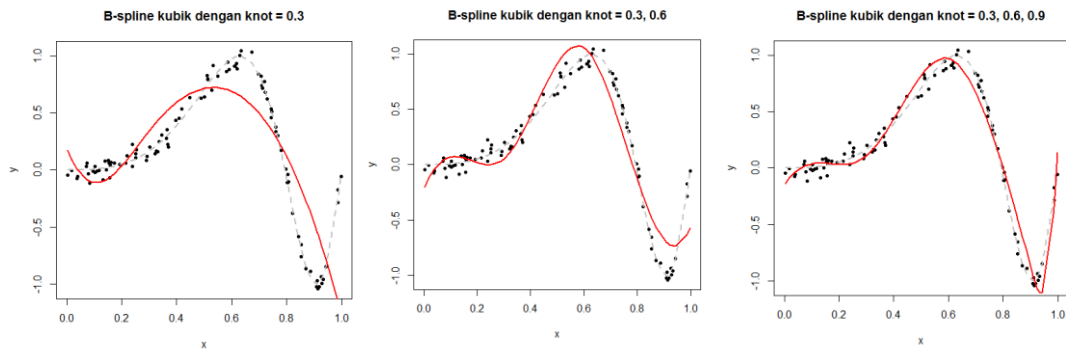
Tabel 18. Taksiran Fungsi B-spline Kubik dengan Knot

$k_1 = 0,3$, $k_2 = 0,6$, dan $k_3 = 0,9$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
$B_{-3,4}$	-0.13715	0.07037	0.05432 .	0.9592	364.7	< 2.2e-16	0.013
$B_{-2,4}$	0.34967	0.12239	0.00528 **				
$B_{-1,4}$	-0.26762	0.08383	0.00193 **				
$B_{0,4}$	1.89691	0.11005	< 2e-16 ***				
$B_{1,4}$	0.09332	0.09287	0.31757				
$B_{2,4}$	1.33909	0.10114	< 2e-16 ***				

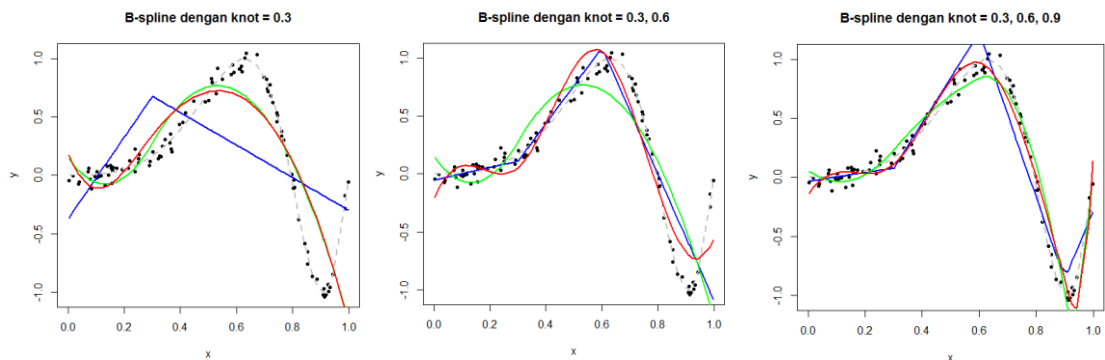
$$B_{3,4} \quad 0.27441 \quad 0.10759 \quad 0.01239 *$$

Kurva taksiran B-spline kubik untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 8 berikut. Tampak bahwa taksiran fungsi B-spline kubik dengan tiga knot lebih mendekati kurva sesungguhnya (garis putus-putus).



Gambar 8. Kurva B-spline Kubik

Secara gabungan, kurva taksiran B-spline linear, kuadratik, dan kubik untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 9. Kurva B-spline kuadratik dan B-spline kubik untuk tiga knot $k_1 = 0,3$; $k_2 = 0,6$; $k_3 = 0,9$ yang paling mendekati kurva sesungguhnya, dengan kata lain kurva taksiran fungsi B-spline kuadratik dan kubik untuk tiga knot $k_1 = 0,3$; $k_2 = 0,6$; $k_3 = 0,9$ mempunyai performa yang paling baik.



Gambar 9. Kurva B-Spline Linear, Kuadratik, dan Kubik

3. *P-spline*

Penggunaan fungsi basis B-spline pada smoothing spline orde empat (kubik) dengan empat knot (dua knot *interior*) akan memberikan enam taksiran koefisien basis. Pada Tabel 19 dapat dilihat, taksiran fungsi P-spline adalah:
 $\hat{f}(x) = -0,164 B_{-3,4}(x) + 0,268 B_{-2,4}(x) - 0,570 B_{-1,4}(x) + 2,516 B_{0,4}(x) - 1,849 B_{1,4}(x) - 0,212 B_{2,4}(x)$ dengan nilai $GCV = 0,011$ dan titik 0.3200673 dan 0.7101114 sebagai knot interior.

Tabel 19. Taksiran Fungsi P-spline dengan empat knot

Basis	Coef	GCV	λ	Knot
$B_{-3,4}$	-	0.01097	3.28E-	0
	0.1642134	3	06	0.320067
$B_{-2,4}$	0.2684155			3
	-			0.710111
$B_{-1,4}$	0.5704224			4
$B_{0,4}$	2.5155371			1
	-			
$B_{1,4}$	1.8485795			
	-			
$B_{2,4}$	0.2121932			

Selanjutnya taksiran fungsi P-spline dengan sepuluh knot (delapan knot *interior*) dapat dilihat pada Tabel 20. Taksiran fungsi P-spline dengan sepuluh knot adalah:

$$\hat{f}(x) = -0,047 B_{-3,4}(x) + 0,000 B_{-2,4}(x) - 0,072 B_{-1,4}(x) + 0,150 B_{0,4}(x) + 0,066 B_{1,4}(x) + 0,599 B_{2,4}(x) + 0,922 B_{3,4}(x) + 1,028 B_{4,4}(x) + 0,121 B_{5,4}(x) - 1,348 B_{6,4}(x) - 0,881 B_{7,4}(x) - 0,024 B_{8,4}(x)$$

dengan nilai $GCV = 0,003$ dan titik-titik 0.1088048, 0.1922796, 0.3200673, 0.4168657, 0.5857644, 0.7101114, 0.7817037, dan 0.887301 sebagai knot interior. Nilai $GCV = 0,003$ yang diperoleh jauh lebih kecil dibanding nilai GCV pada fungsi P-spline dengan empat knot, sehingga performa fungsi P-spline dengan sepuluh knot lebih baik dibanding fungsi P-spline dengan empat knot.

Tabel 20. Taksiran Fungsi P-spline dengan sepuluh knot

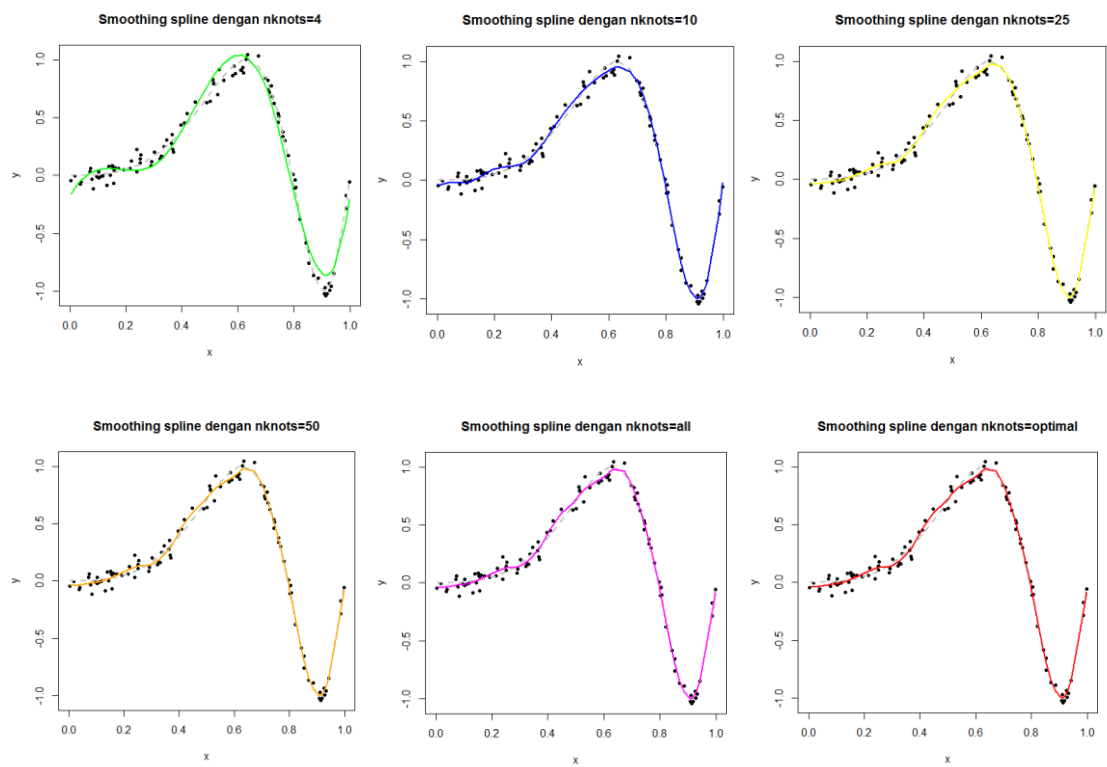
Basis	Coef	GCV	λ	Knot
$B_{-3,4}$	-	0.003223	2.16E-	0
	0.046611899	8	06	0.108804
$B_{-2,4}$	5.19011E-05			8

$B_{-1,4}$	-	0.192279
	0.072232987	6
$B_{0,4}$	0.150176051	0.320067
		3
$B_{1,4}$	0.066388442	0.416865
		7
$B_{2,4}$	0.598560514	0.585764
		4
$B_{3,4}$	0.921724796	0.710111
		4
$B_{4,4}$	1.027582051	0.781703
		7
$B_{5,4}$	0.121211807	0.887301
$B_{6,4}$	-	1
	1.347471269	
$B_{7,4}$	-	
	0.881326613	
$B_{8,4}$	-	
	0.024223202	

Dengan cara yang sama, taksiran fungsi P-spline dengan 25 knot adalah:

$\hat{f}(x) = -0,041 B_{-3,4}(x) - 0,040 B_{-2,4}(x) - 0,024 B_{-1,4}(x) + \dots - 0,438 B_{22,4}(x) - 0,084 B_{23,4}(x)$
 dengan nilai $GCV = 0,003$ dan titik-titik 0.06851225, 0.09290706, 0.11619917, ..., 0.93026352 sebagai knot interior. Nilai $GCV = 0,003$ yang diperoleh tidak berbeda dengan nilai GCV pada fungsi P-spline dengan 10 knot, sehingga performa fungsi P-spline 25 knot sama dengan fungsi P-spline 10 knot. Kurva taksiran P-spline untuk empat knot, sepuluh knot, dan dua puluh lima knot dapat dilihat pada Gambar 10 berikut. Tampak bahwa taksiran fungsi P-spline dengan sepuluh knot dan dua puluh lima knot lebih mendekati kurva sesungguhnya (garis putus-putus). Taksiran fungsi P-spline optimal diperoleh dengan 62 knot yaitu:

$\hat{f}(x) = -0,039 B_{-3,4}(x) - 0,039 B_{-2,4}(x) - 0,038 B_{-1,4}(x) + \dots - 0,084 B_{62,4}(x)$ dengan
 nilai $GCV = 0,003$ dan titik-titik 0.01524163, 0.03716208, ..., 0.98824747 sebagai knot interior.

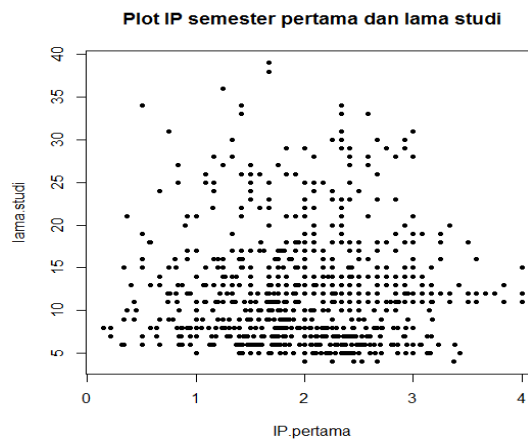


Gambar 10. Kurva P-Spline

B. HASIL TERAPAN

1. Spline dengan Basis Truncated

Berdasarkan data terapan sebanyak 835 mahasiswa Program Studi Statistika FMIPA Universitas Terbuka yang lulus pada masa ujian 1989.2 sampai dengan 2011.2, diperoleh plot Indek Prestasi (IP) pada semester pertama sebagai variabel prediktor dan lama studi sebagai variabel respon seperti pada Gambar 11 berikut.



Gambar 11. Plot IP dan Lama Studi

Selanjutnya, dengan memperhatikan pola hubungan lama studi terhadap Indek Prestasi (IP) pada semester pertama yang sangat acak, akan dicoba untuk menentukan taksiran fungsi spline berorde dua (linear) dengan basis *truncated* pada titik $k_1 = 1$ sebagai knot. Berdasarkan data terapan yang ada, diperoleh taksiran fungsi spline linearnya adalah: $\hat{f}(x) = 11,111 + 1,380 IP - 1,517 (IP - 1)$ dengan nilai $MSE = 39,013$. Baik secara simultan maupun parsial, taksiran parameter yang diperoleh tidak signifikan pada $\alpha = 5\%$ seperti ditunjukkan pada Tabel 21.

Tabel 21. Taksiran Fungsi Spline Linear dengan knot $k_1 = 1$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
Intercept	11.11	1.933	1.26e-08	0.000557	0.231	0.7931	39.013
	1		***	2	9		
IP	1.380	2.084	0.508				
IP-1	-1.517	2.236	0.498				

Jika digunakan titik $k_1 = 1, k_2 = 2$ sebagai knot, maka taksiran fungsi spline linear dengan basis *truncated* adalah:

$$\hat{f}(x) = 10,736 + 2,100 IP - 2,843(IP - 1) + 1,015(IP - 2) \text{ dengan nilai } MSE = 39,025.$$

Baik secara simultan maupun parsial, taksiran parameter yang diperoleh tidak signifikan pada $\alpha = 5\%$ seperti ditunjukkan pada Tabel 22.

Tabel 22. Taksiran Fungsi Spline Linear dengan knot $k_1 = 1, k_2 = 2$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
<i>Intercept</i>	10.736	1.987	8.58e-08***	0.001354	0.3755	0.7707	39.025
<i>IP</i>	2.100	2.264	0.354				
<i>IP-1</i>	-2.843	2.767	0.305				
<i>IP-2</i>	1.015	1.247	0.416				

Selanjutnya jika digunakan titik $k_1 = 1, k_2 = 2, k_3 = 3$ sebagai knot, maka taksiran fungsi spline linear dengan basis *truncated* adalah:

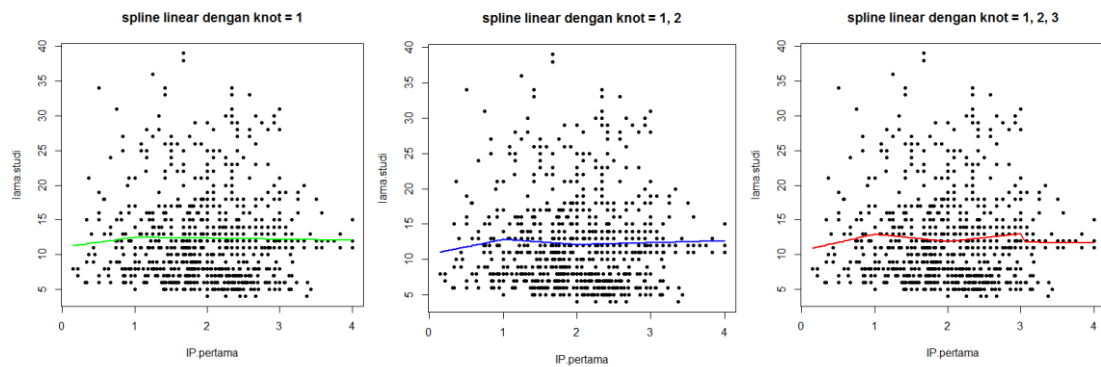
$$\hat{f}(x) = 10,621 + 2,320 IP - 3,386(IP - 1) + 2,181(IP - 2) - 1,213(IP - 3) \text{ dengan nilai } MSE = 38,988.$$

Baik secara simultan maupun parsial, taksiran parameter yang diperoleh tidak signifikan pada $\alpha = 5\%$ seperti ditunjukkan pada Tabel 23.

Tabel 23. Taksiran Fungsi Spline Linear dengan knot $k_1 = 1, k_2 = 2, k_3 = 3$

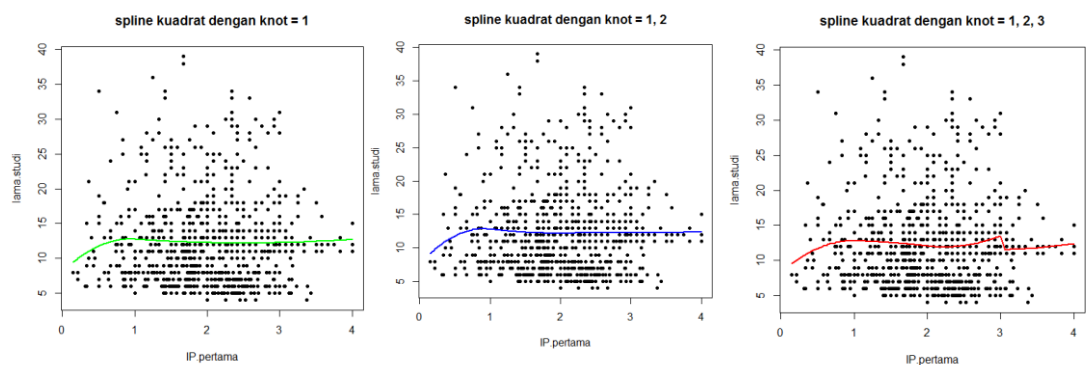
Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
<i>Intercept</i>	10.6214	1.9881	1.18e-07***	0.003459	0.7203	0.5781	38.988
<i>IP</i>	2.3199	2.2692	0.307				
<i>IP-1</i>	-3.3855	2.7963	0.226				
<i>IP-2</i>	2.1812	1.5258	0.153				
<i>IP-3</i>	-1.2126	0.9157	0.186				

Kurva taksiran spline linear untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 12 berikut. Tampak bahwa taksiran fungsi spline linear dengan satu knot, dua knot, dan tiga knot tidak terlalu jauh berbeda satu dengan lainnya, sehingga perlu dicoba spline kuadratik dan kubik.



Gambar 12. Kurva Spline Linear

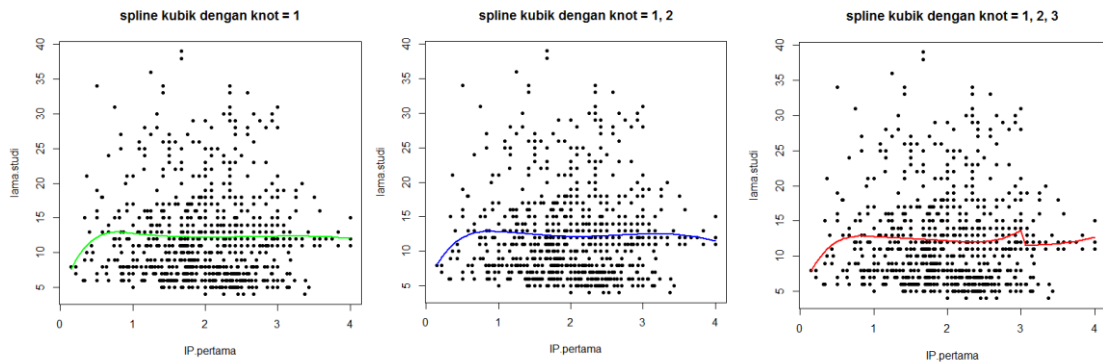
Kurva taksiran spline berorde tiga (kuadratik) dengan basis *truncated* untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 13 berikut. Tampak bahwa taksiran fungsi spline kuadratik dengan satu knot, dua knot, dan tiga knot tidak terlalu jauh berbeda satu dengan lainnya, hal ini didukung nilai *MSE* yang tidak jauh berbeda di antara ke tiga fungsi spline kuadratik. Fungsi spline kuadratik dengan knot $k_1=1$ mempunyai $MSE=39,0$, fungsi spline kuadratik dengan knot $k_1=1, k_2=2$ mempunyai $MSE=39,038$ dan fungsi spline kuadratik dengan knot $k_1=1, k_2=2, k_3=3$ mempunyai $MSE=38,963$.



Gambar 13. Kurva Spline Kuadratik

Kurva taksiran spline berorde empat (kubik) dengan basis *truncated* untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 14. Tampak bahwa taksiran fungsi spline kubik dengan satu knot, dua knot, dan tiga knot tidak terlalu jauh berbeda satu dengan lainnya, hal ini didukung nilai *MSE* yang tidak jauh berbeda di antara ke tiga fungsi spline kubik. Fungsi spline kubik dengan knot $k_1=1$ mempunyai $MSE=39,0$, fungsi spline kubik dengan knot

$k_1 = 1, k_2 = 2$ mempunyai $MSE = 39,038$ dan fungsi spline kubik dengan knot $k_1 = 1, k_2 = 2, k_3 = 3$ mempunyai $MSE = 38,975$.



Gambar 14. Kurva Spline Kubik

2. B-spline

Kurva spline orde dua (linear) dengan fungsi basis B-spline dan satu knot yaitu $k_1 = 1$ mempunyai taksiran kurva :

$\hat{f}(x) = 11,308 B_{-1,2}(x) + 1,183 B_{0,2}(x) + 0,772 B_{1,2}(x)$ dengan $MSE = 39,013$. Secara parsial, taksiran untuk koefisien basis spline linear tidak signifikan pada $\alpha = 5\%$ kecuali koefisien basis $B_{-1,2}$ signifikan seperti ditunjukkan pada Tabel 24.

Tabel 24. Taksiran Fungsi B-Spline Linear dengan knot $k_1 = 1$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
$B_{-1,2}$	11.307	8	1.13e-11	0.00			
		1.642	***	1	0.2319	0.7931	39.013
$B_{0,2}$	1.1827	1.786	0.508				
$B_{1,2}$	0.7722	1.6932	0.648				

Kurva B-spline linear dengan dua knot yaitu $k_1 = 1, k_2 = 2$ mempunyai taksiran $\hat{f}(x) = 11,036 B_{-1,2}(x) + 1,800 B_{0,2}(x) + 1,057 B_{1,2}(x) + 1,601 B_{2,2}(x)$ dengan nilai $MSE = 39,025$. Pada Tabel 25 ditunjukkan bahwa secara parsial, taksiran untuk koefisien basis spline linear tidak signifikan pada $\alpha = 5\%$ kecuali koefisien basis $B_{-1,2}$.

Tabel 25. Taksiran Fungsi B-Spline Linear dengan knot $k_1 = 1, k_2 = 2$

Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
-------	------	----------	----------	----	---	---------	-----

$B_{-1,2}$	11.03 6	1.676	8.13e-11 ***	0.00135 4	0.3755	0.7707	39.025
$B_{0,2}$	1.800	1.941	0.354				
$B_{1,2}$	1.057	1.687	0.531				
$B_{2,2}$	1.601	1.976	0.418				

Kurva B-spline linear dengan dua knot yaitu $k_1=1, k_2=2, k_3=3$ mempunyai taksiran kurva :

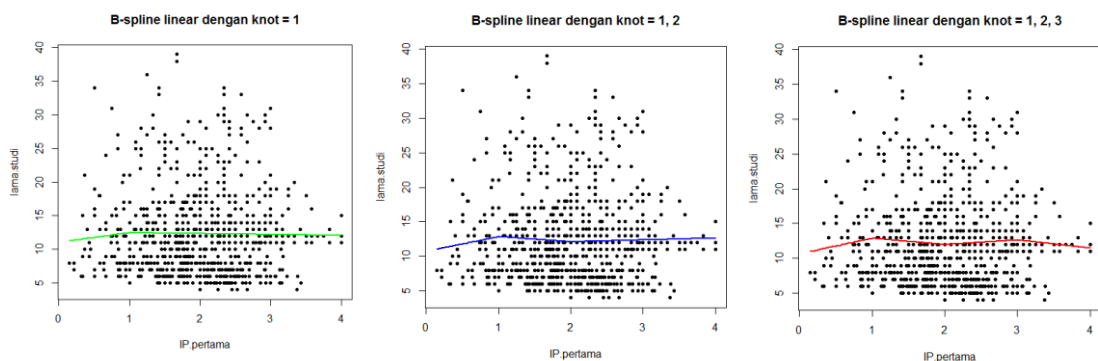
$$\hat{f}(x) = 10,994 B_{-1,2}(x) + 1,896 B_{0,2}(x) + 0,989 B_{1,2}(x) + 1,649 B_{2,2}(x) + 0,529 B_{3,2}(x)$$

dengan nilai $MSE = 39,050$. Secara parsial, taksiran untuk koefisien basis spline linear tidak signifikan pada $\alpha = 5\%$ kecuali koefisien basis $B_{-1,2}$ seperti ditunjukkan pada Tabel 26.

Tabel 26. Taksiran Fungsi B-Spline Linear dengan knot $k_1 = 1, k_2 = 2, k_3 = 3$

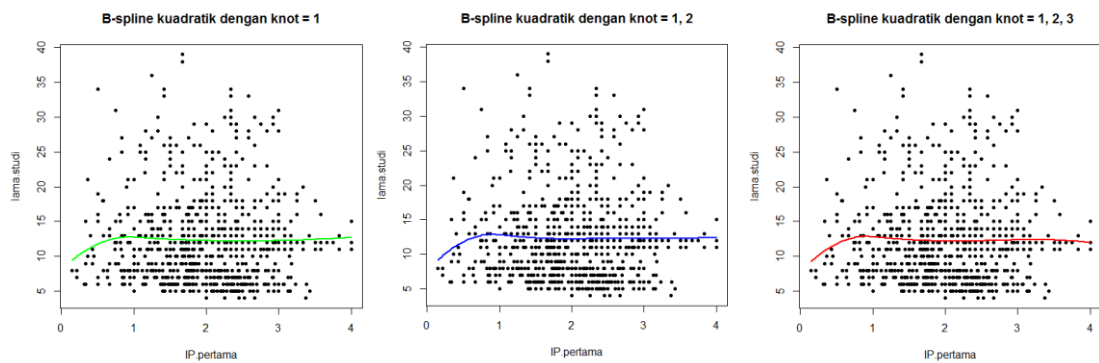
Basis	Coef	SD(coef)	Pr(> t)	R2	F	p-value	MSE
$B_{-1,2}$	10.993 8	1.6778	9.93e-11 ***	0.00193 9	0.4031	0.8065	39.050
$B_{0,2}$	1.8956	1.9461	0.330				
$B_{1,2}$	0.9892	1.6903	0.559				
$B_{2,2}$	1.6492	1.7935	0.358				
$B_{3,2}$	0.5293	2.5037	0.833				

Kurva taksiran B-spline linear untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 15. Tampak bahwa taksiran fungsi B-spline linear untuk ke tiga kurva tidak ada perbedaan, sehingga perlu dicari taksiran fungsi B-spline kuadratik dan kubik.



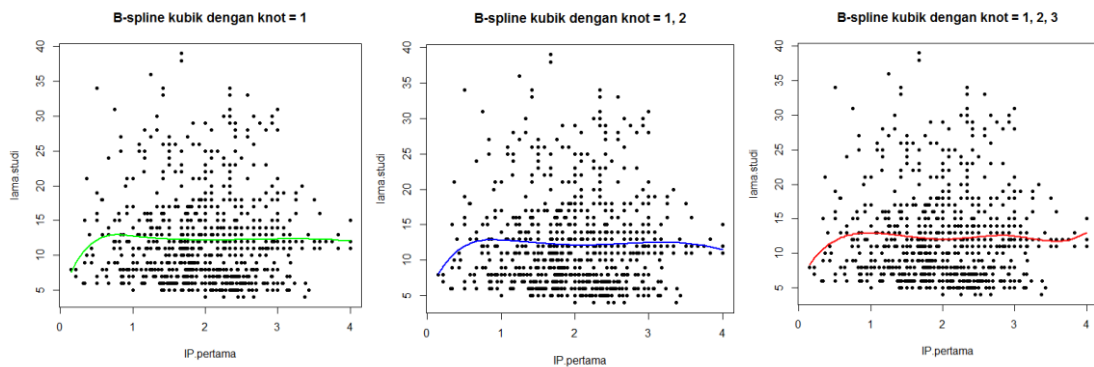
Gambar 15. Kurva B-Spline Linear

Taksiran kurva B-spline berorde tiga (kuadratik) untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 16. Tampak bahwa taksiran fungsi B-spline kuadratik dengan satu knot, dua knot, dan tiga knot tidak terlalu jauh berbeda satu dengan lainnya, hal ini didukung nilai MSE yang tidak jauh berbeda di antara ke tiga fungsi B-spline kuadratik. Fungsi B-spline kuadratik dengan knot $k_1 = 1$ mempunyai $MSE = 39,0$, fungsi B-spline kuadratik dengan knot $k_1 = 1, k_2 = 2$ mempunyai $MSE = 39,038$ dan fungsi B-spline kuadratik dengan knot $k_1 = 1, k_2 = 2, k_3 = 3$ mempunyai $MSE = 39,088$.



Gambar 16. Kurva B-Spline Kuadratik

Kurva B-spline berorde empat (kubik) untuk satu knot, dua knot, dan tiga knot dapat dilihat pada Gambar 17. Tampak bahwa taksiran fungsi B-spline kubik dengan satu knot, dua knot, dan tiga knot tidak terlalu jauh berbeda satu dengan lainnya, hal ini didukung nilai MSE yang tidak jauh berbeda di antara ke tiga fungsi B-spline kubik. Fungsi B-spline kubik dengan knot $k_1 = 1$ mempunyai $MSE = 39,0$, fungsi B-spline kubik dengan knot $k_1 = 1, k_2 = 2$ mempunyai $MSE = 39,038$ dan fungsi B-spline kubik dengan knot $k_1 = 1, k_2 = 2, k_3 = 3$ mempunyai $MSE = 39,063$.



Gambar 17. Kurva B-Spline Kubik

3. *P-spline*

Taksiran smoothing spline dengan basis B-spline kubik pada empat knot (dua knot *interior*) adalah:

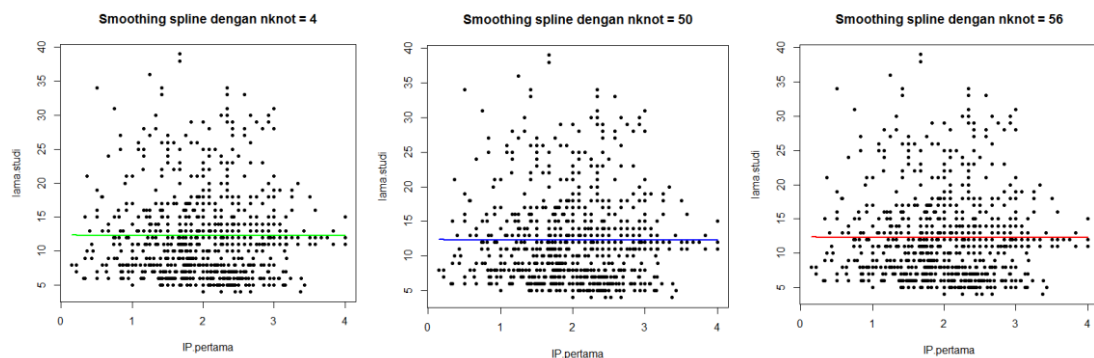
$$\hat{f}(x) = 12,346 B_{-3,4}(x) + 12,338 B_{-2,4}(x) + 12,324 B_{-1,4}(x) + 12,300 B_{0,4}(x) + 12,285 B_{1,4}(x) - 12,274 B_{2,4}(x)$$

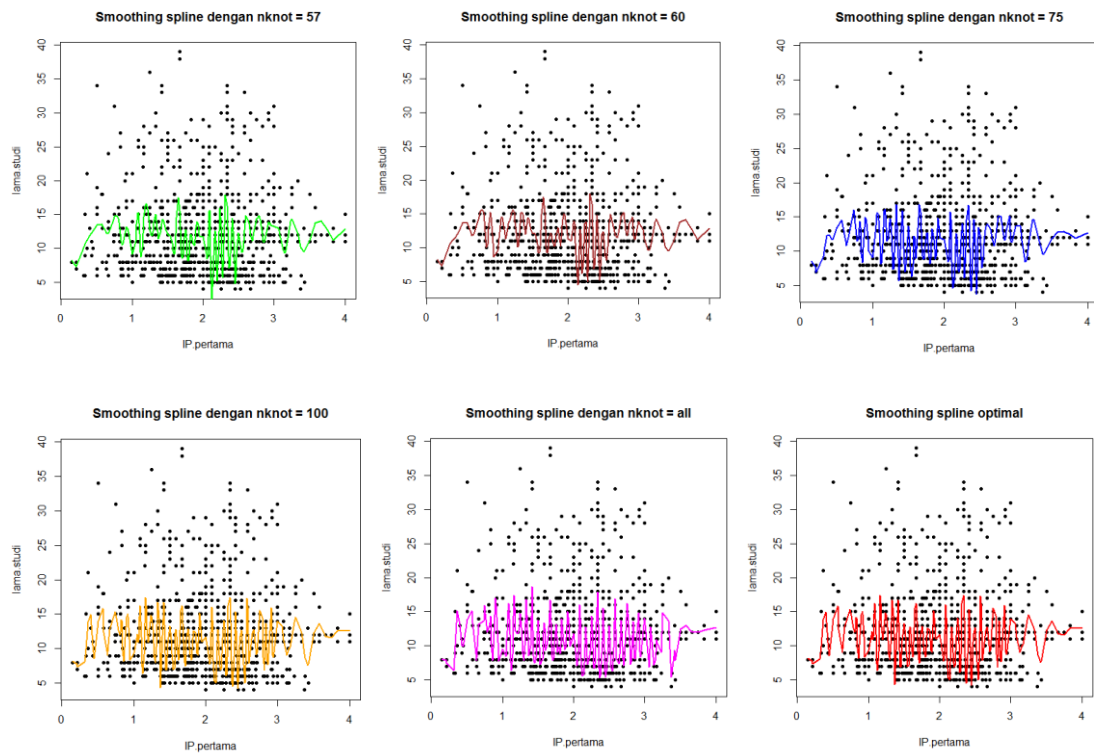
dengan nilai $GCV = 39,077$. Nilai GCV pada smoothing spline untuk berbagai jumlah knot dapat dilihat pada Tabel 27.

Tabel 27. Nilai GCV

<i>nknots</i>	<i>GCV</i>	<i>LAMBDA</i>
4	39.07665	1.77E+07
25	39.07667	44239.25
40	39.07667	5848.219
50	39.07669	2343.159
55	39.07670	1530.198
56	39.07671	1397.238
57	38.55975	1.68E-08
60	38.84518	3.13E-08
75	36.42263	2.50E-08
100	35.04340	1.60E-08
212	33.92993	1.85E-08

Smoothing spline untuk berbagai ukuran knot dapat dilihat pada Gambar 17. Tampak bahwa kurva dengan banyaknya knot sampai dengan 56 masih berbentuk garis lurus, sedangkan untuk banyaknya knot lebih dari 56, kurva tampak sangat fluktuatif. Jika digunakan seluruh knot yaitu banyaknya nilai pengamatan prediktor yang berbeda (ada 212 knot) dalam fungsi, maka diperoleh bentuk kurva yang tidak jauh berbeda dengan smoothing spline dengan 100 knot.





Gambar 17. Kurva P-spline

BAB V

KESIMPULAN DAN SARAN

Berdasarkan hasil pembahasan, diperoleh kesimpulan bahwa:

1. Pada data simulasi,
 - a. pendekatan model spline dengan basis truncated menunjukkan bahwa penduga kurva spline linear, kuadratik, dan kubik pada knot $k_1 = 1$, $k_2 = 2$, $k_3 = 3$ mempunyai performa yang baik dalam memodelkan data .
 - b. pendekatan model spline dengan basis B-spline menunjukkan bahwa penduga kurva B-spline linear, kuadratik, dan kubik pada knot $k_1 = 1$, $k_2 = 2$, $k_3 = 3$ mempunyai performa yang baik dalam memodelkan data.
 - c. pendekatan model smoothing spline dengan basis B-spline menunjukkan bahwa penduga kurva P-spline dengan menggunakan sebanyak 10 knot merupakan model yang mempunyai performa yang baik dalam memodelkan data.
2. Pada data terapan,
 - a. pendekatan model spline dengan basis truncated menunjukkan bahwa penduga kurva spline linear, kuadratik, dan kubik pada berbagai knot mempunyai performa yang kurang baik dalam memodelkan data.
 - b. pendekatan model spline dengan basis B-spline menunjukkan bahwa penduga kurva B-spline linear, kuadratik, dan kubik pada berbagai knot mempunyai performa yang kurang baik dalam memodelkan data.

- c. pendekatan model smoothing spline dengan basis B-spline menunjukkan bahwa penduga kurva P-spline dengan menggunakan berbagai jumlah knot mempunyai performa yang kurang baik dalam memodelkan data.

Adapun saran yang diberikan untuk mendapatkan model yang sesuai untuk data terapan adalah melakukan kajian lanjutan dengan mempertimbangkan lokasi UPBJJ-UT sebagai tambahan peubah prediktor., dengan kata lain perlu dilakukan kajian secara spatial.

DAFTAR PUSTAKA

- Adisantoso, J. 2010. Menentukan parameter pemulus pada model regresi *smoothing spline*. Diunduh pada tanggal 23 Oktober 2012 <http://julio.staff.ipb.ac.id/files/2011/12/tugas3.pdf>
- Botella, O., dan Shariff, K. 2003. B-Spline methods in fluid dynamics. *International Journal of Computational Fluid Dynamics*, 17 (2), 133-149.
- Budiantara, I.N., Suryadi, F., Otok, B.W., Guritno, S. 2006. Pemodelan B-spline dan MARS pada nilai ujian masuk terhadap IPK mahasiswa Jurusan Desain Komunikasi Visual Universitas Kristen Petra Surabaya. *Jurnal Teknik Industri* Vol.8 No.1. UK Petra, Surabaya.
- Eubank, R.L. 1999. *Nonparametric regression and spline smooting*, 2nd ed. Marcel Dekker, New York.
- Hardle, W. 1990. *Applied nonparametric regression*. Cambridge University Press, New York.
- Lee, T.C.M. 2004. *Improved smoothing spline regression by combining estimates of different smoothness*. *Statistics & Probability Letters* 67:133-140. Elsevier.
- Lyche, T., dan Morken, K. 2008. *Spline methods draft*. Department of Informatics for Applications University of Oslo.
- Mood, A.M., Graybill, F.A., dan Boes, D.C. 1974. *Introduction to the theory of statistics*, (3rd Ed.). McGraw-Hill, Japan.
- R Development Core Team . 2011. *R version 2.14.1 : A Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna,

- Austria. ISBN 3-900051-07-0. Diunduh pada tanggal 22 Desember 2011
URL: <http://www.R-project.org/>
- Ruppert, D., Wand, M.P., dan Carroll, R.J. 2003. *Semiparametric regression*. The Press Syndicate of the University of Cambridge, UK.
- Sugiarti, H. & Megawarni, A. 2010. *Koefisien determinasi sebagai ukuran kesesuaian model pada regresi robust*. Makalah Seminar Hasil Penelitian UT tanggal 22 Desember 2010 di LPPM-UT, Jakarta.
- Sugiarti, H. & Megawarni, A. 2011. *Metode least trimmed squared sebagai metode alternatif dalam menaksir koefisien garis regresi*. Makalah Seminar Hasil Penelitian UT tanggal 21 Desember 2011 di LPPM-UT, Jakarta.
- Wibowo, W. 2011. *Metode kuadrat terkecil untuk estimasi kurva regresi semiparametrik spline*. Makalah Seminar Universitas Gadjah Mada Yogyakarta.

BIODATA

Nama : Harmi Sugiarti
Tempat, tanggal lahir : Surabaya, 11 Maret 1967

Pengalaman Kerja

1991 – sekarang : Staf Jurusan Statistika FMIPA-UT
1994 – 1995 : Koordinator Tutorial Tertulis Jurusan Statistika FMIPA-UT
1992 – 1996 : Anggota tim pengajar mata kuliah Statistika di Program Studi Aktuaria dan Perbankan FISIP Universitas Indonesia
1999 – sekarang : Penelaah materi dan format bahan ajar Jurusan Statistika FMIPA-UT
1999 – 2001 : Koordinator Tutorial Tertulis Jurusan Statistika FMIPA-UT
2001 – 2002 : Koordinator Bahan Ujian Jurusan Statistika FMIPA-UT
2002 – 2005 : Anggota Tim Penyunting Majalah Ilmiah BISSOTEK Politeknik Negeri Lhokseumawe
2003 – 2009 : Koordinator Bahan Ajar Jurusan Statistika FMIPA-UT
2007 – 2009 : Sekretaris Jurusan Statistika FMIPA-UT
2009 – 2013 : Ketua Jurusan Statistika FMIPA-UT

Riwayat Pendidikan

1989 : S-1 Statistika Terapan FMIPA Universitas Terbuka
1999 : S-2 Statistika Institut Pertanian Bogor

Pelatihan dan Seminar

1995 : Pelatihan *Basic Statistic Course* di Lembaga Penelitian UT
1998 : Pelatihan Penulisan kisi-kisi soal, penulisan soal, item analysis dan revisi soal di UT
2000 : Pelatihan Transformasi Manajemen Tingkat Eksekutif UT
2001 : *Workshop Loss Distribution* Jurusan Matematika FMIPA UI
2001 : Peserta Seminar Matematika Se-Jakarta & sekitarnya
2001 : Peserta Seminar Nasional Statistika V di Institut Teknologi Sepuluh

- Nopember Surabaya
- 2002 : Peserta Seminar Nasional Matematika di FMIPA Universitas Indonesia
 - 2002 : Pelatihan Penulisan Artikel pada Jurnal Nasional dan Internasional di Universitas Negeri Malang
 - 2003 : Pelatihan *Structural Equation Modelling* di Lembaga Penelitian Universitas Airlangga
 - 2004 : Pelatihan *Sampling* di Lembaga Penelitian Universitas Airlangga
 - 2004 : Seminar Nasional Matematika XII di Universitas Udayana Bali
 - 2008 : Konferensi Nasional Matematika XIV di Universitas Sriwijaya Palembang
 - 2008 : Seminar Nasional FMIPA-UT di Tangerang
 - 2009 : Konferensi Nasional Pendidikan Matematika III di Universitas Negeri Medan
 - 2010 : Konferensi Nasional Matematika XV di Universitas Negeri Manado
 - 2010 : Seminar Nasional FMIPA-UT di Tangerang Selatan
 - 2011 : Seminar Nasional FMIPA-UT di Tangerang Selatan
 - 2012 : Seminar Nasional FMIPA-UT di Tangerang Selatan

Karya Tulis

- 1994 : Malau,R.A., Sugiarti, H. Panduan Praktikum Statistika Program Studi Penyuluh Pertanian
- 1994 : Siregar,H., Malau,R.A., Sugiarti,H. Studi tentang Pengaruh Format Penyampaian Materi Ilmiah melalui Media Noncetak Terhadap Hasil Belajar Mahasiswa FMIPA-UT
- 1998 : Sugiarti,H. Pemilihan Peubah Bebas Terbaik dengan Kriteria Cp untuk Data yang Mengandung *Outlier*
- 1999 : Sugiarti,H. Pemilihan Peubah Bebas Terbaik dengan Kriteria R-Cp untuk Data yang Mengandung *Outlier*
- 1999 : Sugiarti,H. Pemilihan Peubah Bebas Terbaik dengan Kriteria R-Tp untuk Data yang Mengandung *Outlier*
- 2000 : Sugiarti,H., dkk., Analisis Data Statistik, Pusat Penerbitan Universitas Terbuka, Jakarta
- 2001 : Sugiarti,H. Selang Kepercayaan untuk Koefisien Garis Regresi dengan Menggunakan Metode Regresi Robust Jika Terdapat *Outlier*
- 2001 : Sugiarti,H. Kekonsistenan Kriteria RCp dalam Memilih Peubah Bebas Terbaik Jika Terdapat *Outlier*
- 2002 : Sugiarti,H.& Megawarni,. Selang Kepercayaan untuk Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen dengan Metode OLS dan WLS
- 2002 : Sugiarti,H. Kekonsistenan Kriteria RTp dalam Memilih Peubah Bebas Terbaik Jika Terdapat *Outlier*
- 2002 : Sugiarti,H., Herman, Malau, R.A. Analisis Data Statistik
- 2002 : Sugiarti, H. Evaluasi Bahan Ajar Mata Kuliah Model Linear Terapan
- 2003 : Ritonga,H. & Sugiarti,H. Panduan Belajar Statistika untuk PTIK
- 2003 : Sugiarti,H. & Megawarni,A. Penggunaan Metode Regresi Robust untuk mencari Selang Kepercayaan Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen
- 2004 : Sugiarti,H. Tingkat Efisiensi Metode Regresi Robust dalam Menaksir Koefisien Garis Regresi Jika Data Mengandung *Outlier*

- 2004 : Sugiarti,H.& Megawarni,A. Tingkat Efisiensi Metode Regresi Robust dalam Menaksir Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen
- 2006 : Sugiarti,H. Penggunaan Kriteria RCp pada Pemilihan Peubah Bebas Terbaik jika terdapat Multikolinearitas
- 2007 : Soelaeman,I., Pardede,T.,& Sugiarti,H. Pengembangan Materi Modul Analisis Data Statistik
- 2007 : Kusumaningrum,EN., Winarni,I., Sugiarti,H. Pemanfaatan Ekstrak Daun Kemangi (*Ocimum Basilicum*,*LF. Citratum Back*) sebagai Insektisida Nabati untuk pengendalian Hama Ulat *Crocidolomia Binocularis* pada Tanaman Kubis (*Brassica Oleracea*)
- 2008 : Sugiarti,H. Resistensi dan Efisiensi Fungsi Pembobot Huber pada Metode Regresi Robust
- 2008 : Sugiarti,H. Pemilihan Konstanta k pada Metode Regresi Robust
- 2009 : Sugiarti,H. Pola Hubungan Linear antara Partisipasi Mahasiswa dalam Tutorial Online terhadap Nilai Ujian Akhir Semester
- 2010 : Sugiarti,H. Efisiensi Relatif Fungsi Pembobot Huber jika Galat tidak Berdistribusi Normal
- 2010 : Sugiarti,H.& Megawarni,A. Koefisien Determinasi sebagai Ukuran Kesesuaian Model pada Regresi Robust
- 2010 : Sugiarti,H. Penaksir Maximum Likelihood untuk Pola Hubungan Linear antara Partisipasi Mahasiswa dalam Tutorial Online terhadap Nilai Ujian Akhir Semester Mata Kuliah Metode Statistik I (SATS4121)
- 2011 : Sugiarti,H.& Megawarni,A. Metode Least Trimmed Squared sebagai Metode Alternatif dalam Menaksir Koefisien Garis Regresi
- 2012 : Model Pemanfaatan Sumber Daya Alam dan Energi dengan Metode *Least Trimmed Squared*
- 2012 : Konsistensi Koefisien Determinasi sebagai Ukuran Kesesuaian Model pada Regresi Robust

